DOCUMENT RESUME

ED 349 811 FL 020 630

TITLE Focus on Evaluation and Measurement. Proceedings of the National Research Symposium on Limited English Proficient Student Issues (Washington, D.C., September 1991). Volumes 1 and 2.
SPONS AGENCY Office of Bilingual Education and Minority Languages Affairs (ED), Washington, DC.
REPORT NO ED/OBEMLA92-5
PUB DATE Aug 92
NOTE 957p.; For individual papers, see FL 020 631-652.
PUB TYPE Collected Works - Conference Proceedings (021)

EDRS PRICE MF06/PC39 Plus Postage.
DESCRIPTORS Bilingual Education Programs; Elementary Secondary Education; Graduation; Higher Education; Intelligence; Language Research; *Limited English Speaking; Mathematics Education; Referral; Science Education; Scores; Standards; *Student Evaluation; Student Needs; Teacher Certification; *Teacher Education; Testing; Writing Evaluation

ABSTRACT

This document is comprised of a two-volume conference proceedings. The first volume includes the following papers: "Application of Multiple Intelligences: Research in Alternative Assessment" (Joseph Walters) Discussants: Vera John-Steiner, Sue Teele; "Improving Bilingual Education Programs through Evaluation" (Alan L. Ginsburg); "Language Testing Research: Lessons Applied to LEP Students and Programs" (John W. Oller, Jr.) Discussants: Fred Davidson, Myriam Met; "Performance Assessment of Language Minority Students" (Jack S. Damico) Discussants: J. Michael O'Malley, Cecilia J. Navarrete; "SEA Usage of Alternative Assessment: The Connecticut Experience" (Joan Boykoff Baron) Discussants: Mary Jean Habermann, Richard A. Figueroa; "Portfolio Assessment and LEP Students" (Russell L. French) Discussants: Alice J. Kawakami, Daniel Koretz; "A Political/Sociological Critique of Teacher Education Reforms: Evaluation of the Relation of Power and Knowledge" (Thomas S. Popkewitz); "Assessing Appropriate and Inappropriate Referral Systems for LEP Special Education Students" (Alba A. Ortiz) Discussants: Ann C. Willig, Sherry R. Migdail; "The Assessment of Alternative Certification Practices" Panelists: Annalisa Allegro, Migdalia Romero, Elena Izquierdo, Discussant: Barbara Clements; "Teachers for Language Minority Students: Evaluating Professional Standards" (Eugene Garcia); "Evaluating LEP Teacher Training and In-Service Programs" (Stephanie Dalton, Ellen Moir) Discussants: Lynn Malarz, Victoria Jew. Volume II includes the following papers: "Issues in Policy, Assessment, and Equity" (Eva L. Baker) Discussants: Lorraine Valdez Pierce, Peter M. Byron; "Testing LEP Students for Minimum Competency and High School Graduation" (Kurt F. Geisinger) Discussants: Michele R. Hewlett-Gomez, Lawrence M. Rudner; "Innovative Practices in the Identification of LEP Students" (JoAnn Canales) Discussants: Julia Lara, Robert Rueda; "Test Score Pollution: Implications for LEP Students" (Thomas Haladyna) Discussants: Gary Hargett, Maria Pennock-Roman; "LEA Title VII Program Evaluations" Panelists: Raj Balu, Jesus Salazar, Tomi D. Berney, Discussant: Robert Martinez; "Evaluating Mathematics Education of LEP Students in a Time of Educational Change" (Walter Secada) Discussants: Penelope L. Peterson, Mary Lindquist; "Science

Education as a Sense-Making Practice: Implications for Assessment" (Beth Warren and Ann S. Rosebery) Discussants: Ron Rohac, Sau-Lim Tsang; "Holistic Writing Assessment of LEP Students" (Liz Hamp-Lyons) Discussants: Denise McKeon, Joy Kreeft Peyton; "A Superintendent's Evaluation of Teacher Education Reforms" (Peter J. Negroni); "Designing an IHE Teacher Training Program for Specific LEP Student Instructional Needs" (John E. Steffens) Discussants: Virginia Collier, Rosita G. Galang; and "Educational Research and Teacher Training for Successfully Teaching LEP Students" (Carl A. Grant) Discussants: Margarita Calderon, Li-Rong Lilly Cheng. (VWL)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

ED349811

Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues:

FOCUS ON EVALUATION AND MEASUREMENT

VOLUME 1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

United States Department of Education
Office of Bilingual Education and Minority Languages Affairs

BEST COPY AVAILABLE

Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues:

FOCUS ON EVALUATION AND MEASUREMENT

Washington, D.C.
September 1991

VOLUME 1

United States Department of Education
Office of Bilingual Education and Minority Languages Affairs

Published August 1992

"The papers in this publication were commissioned by the Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education. However, these contents do not necessarily represent the policy of the U.S. Department of Education, and endorsement by the Federal Government should not be assumed."

FOREWORD

With this publication OBEMLA adds twenty research papers to the ten presented at the First National Research Symposium in 1990. The focus of these papers, delivered at the Second Symposium on LEP Student Issues, is especially timely. Evaluation -- understood not only as a technique but, more important, as a habit of thought -- is still in its infancy. This is as true of education as it is of business or social services. One has merely to read the daily newspaper regularly to become aware that evaluation is a recurring preoccupation in any institution -- whether a Fortune 500 corporation or a private academy or a drug rehabilitation center -- that convenes people around a shared task. Evaluation enables us to discover certain facts about the past and the present -- what works and what does not. But that is not enough. Evaluation must also reveal to us the how's and why's so that we can make judgments about the future, so that we can deliberately choose our next steps.

At last year's symposium I noted the importance of research, from which I expect both the theoretical framework and the factual grounding of effective second language learning processes. To this affirmation, I want to add another: ultimately, the conclusions of research must be accessible to the people who make policy, who teach, who design curricula, and, yes, even to the people who seem the furthest removed from academia -- the plain ordinary parents of plain ordinary language minority students. My words are not intended to bash "pure" scholarship or "ivory towers"; above all, they do not dismiss those who study and think and analyze and construct new theory. On the contrary, I respect the work of scholars and value their contribution to a task that is large enough to utilize the diverse talents of all of us. But I do mean to underline a central fact: if the knowledge and the understanding created by research do not ultimately enlighten the publics I mentioned, the field will never reach the breakthrough insights and decisions demanded by the mammoth needs of students. In relation to the topic at hand, evaluation, the broad accessibility of research findings is key to the educational regeneration we seek. I challenge the research community, therefore, to be inventive about the interpretation and transmission of findings.

I know that the research reported in these papers will make significant contributions to the thousands who work with and for language minority students. We at OBEMLA will surely take them to heart. I am proud of OBEMLA's role in promoting research and grateful to Dr. Carmen Simich-Dudgeon and her staff for their efforts in planning and conducting the symposia.

Rita Esquivel
Director
Office of Bilingual Education
and Minority Languages Affairs
U.S. Department of Education

(Note: On May 30, 1992, Nguyen Ngoc-Bich assumed the role of OBEMLA's Acting Director. Rita Esquivel resigned her position as OBEMLA's Director to resume her career with the Santa Monica-Malibu Unified School District, California.)

6

ACKNOWLEDGEMENTS

The OBEMLA Research and Evaluation Staff (RES) wishes to express its appreciation to the authors of the commissioned papers that appear in this volume. These papers were presented at the Second Research Symposium on Limited English Proficient (LEP) Issues held in Washington DC, September 4-6, 1991. The authors are: Joseph Walters, Harvard University; Alan Ginsburg, United States Department of Education; John Oller, University of New Mexico, Albuquerque; Jack Damico, University of Southern Louisiana; Joan Baron, Connecticut State Department of Education; Russell French, University of Tennessee; Thomas Popkewitz, University of Wisconsin, Madison; Alba Ortiz, University of Texas, Austin; Eugene Garcia, University of California, Santa Cruz; and Stephanie Dalton and Ellen Moir, University of California, Santa Cruz.

The collection of papers in this volume represents the United States Department of Education's continued commitment to provide a forum for the exchange of information and discussion of diverse points of view regarding the education of LEP students.

RES also wishes to recognize and thank the following Symposium discussants and panelists: Sue Teele, University of California, Riverside; Vera John-Steiner, University of New Mexico; Fred Davidson, University of Illinois, Urbana/Champaign; Miriam Met, Montgomery County Public Schools, Maryland; Michael O'Malley, Georgetown University; Cecilia Navarrete, University of New Mexico; Mary Jean Habermann, New Mexico Department of Education; Richard Figueroa, University of California, Davis; Alice Kawakami, Regional Educational Laboratory, Hawaii; Daniel Koretz, RAND Corporation; Ann Willig, Florida Atlantic University; Sherry Migdail, Consultant; Annalisa Allegro, New Jersey Department of Education; Migdalia Romero, Hunter College; Elena Izquierdo, District of Columbia Public Schools; Barbara Clements, Council of Chief State School Officers; Lynn Malarz, Association for Supervision and Curriculum Development; and Victoria Jew, California State University.

RES is especially appreciative of the collaboration provided by the U.S. Department of Education's Office of Educational Research and Improvement (OERI), and particularly by Dr. René Gonzalez, OERI's Center Monitor for the National Center for Research on Cultural Diversity and Language Learning at the University of California, Santa Cruz.

Finally, my special thanks to Alex Stein, Cindy Towsner, and Cecile Kreins of my staff, for assisting authors, discussants, and panelists during the paper submission process. My sincere appreciation to Timothy D'Emilio and Sharon Coleman of my staff for their hard work in coordinating the publication of Volume I and Volume II, the proceedings of the Second Research Symposium with a focus on evaluation and measurement.

Carmen Simich-Dudgeon
Director, Research and Evaluation
Office of Bilingual Education and Minority Languages Affairs
US Department of Education

CONTENTS

Volume I

FOREWORD iii

ACKNOWLEDGEMENTS v

INTRODUCTION xi

Application of Multiple Intelligences Research in Alternative Assessment 1

Joseph Walters

Discussants:

Vera John-Steiner 19
Sue Teele 23

Improving Bilingual Education Programs Through Evaluation 31

Alan L. Ginsburg

Language Testing Research: Lessons Applied to LEP Students and Programs 43

John W. Oller, Jr.

Discussants:

Fred Davidson 125
Myriam Met 131

Performance Assessment of Language Minority Students 137

Jack S. Damico

Discussants:

J. Michael O'Malley 173
Cecilia J. Navarrete 183

SEA Usage of Alternative Assessment: The Connecticut Experience 187

Joan Boykoff Baron

Discussants:

Mary Jean Habermann 235
Richard A. Figueroa 243

Portfolio Assessment and LEP Students 249

Russell L. French

Discussants:

Alice J. Kawakami 273
Daniel Koretz 281

A Political/Sociological Critique of Teacher Education Reforms: Evaluation of the Relation of Power and Knowledge 287

Thomas S. Popkewitz

Assessing Appropriate and Inappropriate Referral Systems for LEP Special Education Students 315

Alba A. Ortiz

Discussants:

Ann C. Willig 343
Sherry R. Migdail 353

The Assessment of Alternative Certification Practices

Panelists:

Annalisa Allegro 363
Migdalia Romero 367
Elena Izquierdo 373

Discussant:

Barbara Clements 377

Teachers for Language Minority Students: Evaluating Professional Standards **383**

Eugene Garcia

Evaluating LEP Teacher Training and In-Service Programs **415**

Stephanie Dalton and Ellen Moir

Discussants:

Lynn Malarz 447
Victoria Jew 453

INDEX **457**

INTRODUCTION

This is Volume I of two volumes that contain the proceedings of the Second National Research Symposium on Limited English Proficient Student Issues. The Symposium represented a collaborative effort between the Office of Bilingual Education and Minority Languages Affairs (OBEMLA) and the Office of Educational Research and Improvement (OERI) and was held in Washington DC, September 4 through 6, 1991.

The general theme of the papers in these volumes is evaluation and measurement. It is an effort on the part of OBEMLA to promote the dissemination of state-of-the-art information regarding key issues in the education of school-age students of limited English proficiency (LEP). Specifically, the papers discuss both the theory and its application in the area of educational evaluation and measurement and the role of assessment in terms of accountability and program improvement at the federal, state, and local levels. In addition, evaluation and measurement issues in other areas are discussed. For example, the evaluation of teacher education programs, both at the preservice and in-service levels and the evaluation of curricula, e.g., science and math, in view of advances in these and other fields are topics covered in these volumes. Other topics, including the applications of foreign language testing to second language learning are discussed,as is research on multiple intelligence, its present and future impact on changes in the way we envision the field of evaluation and measurement,and its initial applications to LEP student and program evaluation.

We believe that dissemination of innovations in evaluation and measurement are at the core of the school reform movement. The papers in this volume we hope will act as catalysts to dialogue between practitioners and researchers about alternative assessment theories, methods, and strategies, and their potential application to the assessment of LEP students' language and subject matter knowledge. Furthermore, we hope that discussion will expand to include issues of program evaluation and improvement. Alternative assessment practices, including portfolio assessment and holistic writing assessment, are innovative trends in evaluation and measurement whose time has come. We encourage further study of these innovations and discussion of diverse points of view on the merits and constraints of these methods in the education of LEP students.

The remainder of this section consists of brief summaries of the main issues discussed in each paper.

In his paper "Application of Multiple Intelligences Research in Alternative Assessment," Joseph Walters explores the implications of the theory of multiple intelligences for education in general and for the education of children of limited English proficiency in particular.

The author introduces a theoretical treatment of the concept of intelligence that provides for human intellectual diversity, and contrasts this view with the more traditional notion of intelligence. Dr. Walters draws several implications for education from this theory, paying particular attention to the question of assessment and trying to show why this view of intelligence forces us to rethink some of the fundamental assumptions we hold about the assessment of learning. Finally, Dr. Walters suggests implications for bilingual and multicultural learning.

In "Foreign Language Testing: Lessons Applied to LEP Students," John Oller proposes a theory of human representational abilities and makes recommendations for testing and teaching LEP students and evaluating the programs that purport to serve them. Discourse-based tasks grounded in actual language performances in real life contexts are recommended rather than surface-oriented procedures that focus on bits and pieces of language. Dr. Oller suggests that the full range of students' semiotic abilities should be taken into consideration and they should be tested in their native languages and observed in a broad range of contexts.

In "Performance Assessment of Language Minority Students" Jack Damico writes about the characteristics necessary for successful performance assessment and the assessment process. Dr. Damico suggests that performance assessment of language minority students requires the application of theoretically defensible procedures that are carefully designed and systematically implemented. Due to the differences between language minority students in the schools and those students in English as a Second language (ESL) or English as a Foreign language classes typically studied by language testing researchers, performance assessment in the schools must involve the utilization of procedures that are highly authentic, more functional, more descriptive, and more individualized than those typically recommended by second language testing researchers. Dr. Damico's paper proposes a descriptive approach to performance assessment that is theoretically defensible and psychometrically sufficient.

Joan Boykoff Baron's paper, "SEA Usage of Alternative Assessment: The Connecticut Experience," suggest that there is growing dissatisfaction with current over-reliance of schools on multiple-choice tests. The five sections of this paper describe Connecticut's attempts over the past decade to develop assessments which use meaningful performance tasks to determine what students know and can do.

The first part of the paper describes the Connecticut Assessment of Education Progress program which, between 1982 and 1987, successfully used performance assessments to assess what students know and can do in art and music, business and office education, drafting, English language arts, graphic arts, foreign language, science, and small engines. Sample exercises and their scoring rubrics are presented and described. The second part of the paper describes the Connecticut Mastery Testing Program which, since 1985, has included the use of calculators for mathematics problem-solving in Grade 8, and the use of writing samples and note-taking exercises in Grades 4, 6, and 8. In the third part of her paper, Dr. Baron describes the work that resulted from Connecticut's receipt of a grant from the National Science Foundation. The fourth part synthesizes the characteristics of effective performance tasks and sets forth some of the advantages of using performance assessments to determine what students know and do. The paper concludes by discussing some of the issues inherent in using performance assessments with students of limited English proficiency.

In "Portfolio Assessment," Russell French states that the development of authentic assessments for all students is one of the major educational issues of the 90s. He argues that students' ability to function in our complex world cannot be measured with current standardized tests and that current tests do not measure what students truly know and are able to do.

This paper presents the arguments for alternative assessments and the cautions to be exercised as the authentic assessment movement gathers momentum. It then examines current work in authentic assessment and the assessment needs of LEP students. After defining and differentiating among the three most widely utilized performance assessment methodologies (performance tasks, exhibitions and portfolios), the author focuses attention on portfolio contents, utilization, and design issues. Finally, he identifies a series of questions and discussion items which can be used by educators attempting to design student assessment portfolios.

Thomas Popkewitz discusses issues of teacher education evaluation within a broad socio-historical perspective in his paper, "A Political/Sociological Critique of Teacher Education Reforms." Dr. Popkewitz posits that evaluation is a state strategy to produce social amelioration. Its categories and distinctions often redefine social issues into administrative categories that can be ordered, supervised, and controlled.

Dr. Popkewitz goes on to discuss the role of evaluation in society. He suggests this role is to help to illuminate the tensions, contradictions, and ambiguities that underlie the realization of educational reform. The author concludes by suggesting that the reform priorities of schools are indelibly tied to social, cultural, and economic conditions; these cannot be lost in the methodologies of evaluation.

Alba Ortiz is the author of "Assessing Appropriate and Inappropriate Referral Systems for LEP Special Education Students." In her paper, she focuses on the lack of educational progress of Hispanics and other language minority students in special education as these students are likely to be referred for special services because of academic difficulties.

According to Dr. Ortiz, there is evidence to suggest that language minorities are over represented in programs for the learning disabled and, with the exception of Asian students, under represented in programs for the gifted and talented. More minorities continue to be served in special education than would be expected from their percentage of the general school population. With projections that one of every three Americans in this country will be black, brown, or Asian by the year 2000, the author suggests that greater attention must be given to assuring that multicultural populations experience success in mainstream education and, if referred to special education, that procedures used to assess functioning levels and to recommend services reflect that those involved in the decision-making process understand how language and culture influence performance.

In his paper, "Evaluating Credentialing Programs for Teachers of LEP Students," Eugene Garcia suggests ways in which to enhance the educational plight of LEP students by focusing on the education professionals who directly serve these students on a daily basis.

The author suggests that political debate regarding the education of language minority students has centered on the instructional use of the native and/or the English language as a medium and/or target of instruction. However, educational professionals and researchers recognize that the more specific concerns have become identification, implementation, and evaluation of effective instruction of the ethnolinguistic minority student population.

While the central theme of Dr. Garcia's paper remains on the instructional role of the native language, the discussion focuses on how approaches to this instructional role affect the type of teachers who serve these students. A major presupposition of this discussion is that "who" does the teaching is of major significance regardless of the language minority educational model which is being implemented.

The paper, "Evaluating LEP Teacher Training and In-Service Programs," co-authored by Stephanie Dalton and Ellen Moir is among the least reported issues in the literature of teacher education research. The consequences of this neglect says Dr. Dalton and Ms. Moir are evident not only in program evaluation's underdevelopment and in unexamined teacher education programs but also in the individual experiences of increasing numbers of teachers nationwide. The authors state that LEP teacher training and in-service programs can provide teachers with the assistance necessary to increase the academic performance of linguistically and culturally diverse students.

In this paper, the authors first summarize the history of teacher education evaluation, particularly its methodology, and then examine content recommendations coming from current research on effective education of linguistically diverse students. Secondly, they report their experiences with two evaluated teacher education programs, a University of Hawaii alternative program titled Pre-Service Education for Teachers of Minorities (PETOM) and the California New Teacher Project (CNTP) at the University of California at Santa Cruz. Based on the presentation of teacher education program evaluation literature, the findings of recent research on effective teaching and learning models for linguistic minorities, and the experiences of two programs, the authors conclude with recommendations for LEP preservice and in-service teacher education program evaluation.

Carmen Simich-Dudgeon
Director, Research and Evaluation
Office of Bilingual Education and Minority Languages Affairs
US Department of Education

Application of Multiple Intelligences Research in Alternative Assessment

Joseph Walters
Harvard University

Introduction

Like many urban areas, Harvard Square in Cambridge, Massachusetts, contains a number of restaurants that open onto sidewalks and public spaces. On a pleasant Sunday morning, one such area, near the center of the Square, is filled with people who have come together to talk, play games, and read. Clusters of players and spectators are engrossed in games of chess and backgammon; a musician plays an amplified guitar; a juggler performs in an open space; the Times crossword puzzle is the subject of debate at one table; and, of course, the entire area is filled with animated conversation.

What struck the author about this scene, especially as he was gathering ideas for this paper, was the diversity of the human skills on display in this small space. As he looked about, he could easily pick out a variety of pursuits and challenges -- the games of chess and backgammon, word puzzles, musical and kinesthetic performances, social interaction, and so on. And yet, nothing in this scene was unusual. The diversity that he was seeing was completely familiar.

Another striking feature of this scene was how much of it builds on problem solving. Games like chess and backgammon allow the players to pose problems for one another. Puzzles are taken up as a challenge posed by the puzzle's author. Performances in music and movement require the solution of problems of a different sort.

This scene was a reminder of the need that humans have to create challenges and pose problems as a form of recreation. What's more, there is an inevitable variety to the nature of those challenges. For one person, chess is a fascinating and fulfilling game, while for a second person chess is impenetrable, a foreign language. The crossword puzzle for these two people may appeal in just the opposite manner.

What is it about humans that yields this intellectual diversity? And how is this diversity reflected in learning? In this paper, the author will introduce a theoretical treatment of the concept of intelligence that provides for this diversity and will contrast this view with the more traditional notion of intelligence. Next, he will draw from this theory several implications for education, paying particular attention to the question of assessment. He will try to show why this view of intelligence forces us to rethink some of the fundamental assumptions we hold about the assessment of learning. Finally, he will draw from the discussion of "multiple intelligences" and assessment a consideration of several specific implications for bilingual and multicultural learning.

The Question of Intelligence

To begin, the author defines the term intelligence as an individual's ability to solve problems or fashion products. In the traditional view -- one held by many psychologists -- intelligence is a human trait that varies from one individual to the next such that the individual with a great deal of this trait (the more intelligent individual) is more adept at solving problems and fashioning products. Indeed, it doesn't matter what the problem is. For any problem the highly intelligent person will be more likely to solve it than the less intelligent person.

To examine or test this trait in individuals, psychologists have constructed a large set of test problems and asked people to solve them. From the solutions offered to these test problems (some individuals solve these problems more accurately, quickly, insightfully, and so on) the psychologists predict which individuals will be most likely to solve any problem accurately and insightfully. In fact, the actual problems on the test aren't of particular interest and they are often quite trivial. "Who wrote The Iliad?" Or, "Recite these digits backwards, 2,5,3,4,7." Questions like these do not in themselves pose interesting problems but the psychologists use them to identify those individuals who are most effective problem solvers. Since there is a single trait of intelligence, these tests, with their rather trivial questions, identify all individuals who are well endowed with that trait. Psychologists then predict that those highly-endowed individuals will most likely display intelligent behavior in the future.

This traditional view of intelligence as a singular trait presents us with a difficulty. When we try to apply it to human behavior in the world, we find that many people who display particular talents and proclivities do not "test well" on our measures of intelligence. For example, in the Harvard Square street scene, we may find that the backgammon player can answer certain questions on the IQ test quite accurately but has trouble with others; the musician displays a very different pattern of answers. In other words, we can identify talented individuals in the world, but we do not find that the trait of intelligence, as revealed by intelligences tests, has much to do with *these talents. Indeed, when we look at the variety of things that* people can do, we begin to think that there might be more to "intelligence."

We are left with this problem: We recognize "intelligence" as an important construct in understanding how humans learn and solve problems, but the traditional view of intelligence and the tests that have been designed to appraise it are too limited in scope. Human performance appears to be too complex and diverse to be captured in this single dimension. What we are left looking for, then, is a theory of intelligence that can reflect the complexity of skills and performances that humans exhibit in the world. By examining those skills, we might reason backwards to the "intelligences" that must be responsible.

The Theory of Multiple Intelligences

The theory of Multiple Intelligences (MI) takes this perspective as its starting point. Developed by Howard Gardner and described in his book Frames of Mind (1983), the theory posits seven distinct and universal capacities. These capacities, or intelligences, are innately endowed in all humans; but at the same time, they are manifested quite differently in different cultures. For example, the linguistic intelligence, an innate and universal capacity found in all societies, can appear through writing in one culture, public speaking in a second, and a secret anagrammatic code in a third. Or the spatial intelligence, another ability found in all societies, is displayed in many different ways, from navigation, to the game of chess, to the science of geometry. So, the intelligences are innate and universal, but they are distinctly shaped by the cultures they appear in.

To be useful, the capacities that we identify must be relatively few in number. A theory with too many capacities that were too finely sliced would be less interesting theoretically and much less useful to practitioners. The candidate capacities, to be certified as intelligences, must also be established as distinct and independent on empirical grounds. For example, we know from studies of brain damage, that the linguistic capacity can be damaged while other cognitive functions remain unchanged; this indicates that the linguistic function is separate from those other functions. Studies of idiot savants, who display one skill at a sophisticated level and yet are well below normal in other areas, again help identify distinct cognitive functions. Research from child development, child prodigies, cross cultural investigations, as well as the traditional research of psychological training studies and psychometric research complete the empirical criteria that are applied to candidate skills. Only those faculties for which there is reasonably strong evidence are included in the list of multiple intelligences.

Seven faculties survive this test. *Next, I will examine these seven* and make several observations about each one.

Linguistic Intelligence

Although it is easy to accept the idea that linguistic skill is an intelligence -- almost all tests of intelligence contain items that reveal this faculty -- we also find evidence from our various sources to include it. For one thing, there is a very specific region of the brain, "Broca's Area," that is responsible for interpreting linguistic information. Also, stroke victims reveal a loss of the linguistic faculty while other cognitive processes remain unchanged. A person with damage to Broca's Area can understand words but cannot assemble these components into anything other than the simplest sentences.

We can also find examples of child prodigies in the linguistic realm. For example, T.S. Eliot, at the age of ten during his winter vacation, created his own magazine, which he called "Fireside." There were eight issues and each issue contained poems, adventure stories, humor, recipes, and a gossip column. When examined this material displays the talent of this budding poet and critic (Soldo, 1982).

The gift of language is found in all populations and in all cultures. It develops according to a very predictable schedule in infants. For these reasons, the linguistic faculty passes the empirical test to be included in our list of intelligences.

Logical-mathematical Intelligence

Logical and mathematical abilities, like the linguistic skill, are often associated with the term intelligence; again, many items on tests of intelligence tap these abilities directly. However, the logical-mathematical aptitude must also be included on our list because it passes the empirical test that we have established for multiple intelligences.

The logical-mathematical ability is distinct from the linguistic ability and often the mind solves logical problems without putting them into words. An example comes from the biography of Barbara McClintock, Nobel laureate in genetics. McClintock studied maize and one day her field results, literally taken in a corn field, indicated a pollen sterility different from that predicted by the prevailing theory. McClintock returned to her office and thought about the problem for a while. Suddenly the solution came to her. She ran back to the corn field, announced her solution to her skeptical colleagues, and then sat down in the field and sketched out a proof on a paper bag.

> I worked out the solution, step by step, and I came out with [the same result]. [They] looked at the material and it was exactly as I had said it was; it worked out exactly as I had diagrammed it. Now, why did I know, without having done it on paper? Why was I so sure? (Keller, 1982, p. 104)

This story reminds us that the mathematical ability is distinct from the linguistic skill. It also shows the speed with which talented individuals can develop solutions to mathematical problems.

Spatial Intelligence

Like linguistic and logical-mathematical abilities, the spatial skill appears on numerous tests of intelligent behavior. The Wechsler Intelligence Scale for Children (WISC), for example, includes a subscore that measures spatial abilities through tasks that ask the subject to visualize objects in a rotated configuration.

The spatial intelligence is brought to bear in a variety of activities from solving geometry problems, playing chess, navigating a boat, or reading a map. Evidence from brain research, child development, and anthropological accounts support its inclusion on our list. For example, consider the spatial skills of sailors in the Caroline Islands in the South Seas:

> Navigation around the Caroline Islands is accomplished without instruments. The position of the stars, the weather patterns, and water color are the only sign posts. Each journey is broken into a series of segments. During the actual trip the navigator must envision mentally a reference island as it passes under a particular star and from that he computes the number of segments completed, the proportion of the trip remaining, and any corrections in heading that are required. The navigator cannot see the islands as he sails along; instead he maps their locations in his mental "picture" of the journey. (Gardner, 1983)

These various uses of the spatial intelligence remind us that although the intelligences are innate and universal, they appear in very different contexts from one culture to another. Also, spatial intelligence in the blind population underscores the important difference between the intelligence (the spatial ability) and the various modalities of sense data (seeing and touching). A blind person is perfectly competent spatially, creating mental maps of an environment or recognizing objects by touch, without receiving the visual data that are so important to spatial judgments for the seeing person.

Musical Intelligence

The biographies of famous musicians, like those of mathematicians, contain many stories of the early emergence of extraordinary talent at an early age, even before the child has received musical training. For example, at the age of 3, Arthur Rubinstein was taken to the great teacher and violist, Jacob Joachim, because his parents, who themselves lacked musical training, recognized his extraordinary talent. In this interview, young Arthur was asked to call out chords struck on the piano, to play a theme from a Schubert symphony after Joachim had hummed it, and to add the correct harmonies to the phrase and to transpose it. Joachim concluded from this brief interaction: "This boy may become a great musician... he certainly has the talent for it. Let him hear some good singing, but do not force music on him. When the time comes for serious study, bring him to me and I shall be glad to supervise his artistic education." (Rubinstein, 1978). Of course, Joachim was correct in his assessment and Rubinstein returned to Berlin to study with Joachim five years later.

Our review of the empirical evidence, including biographies of child prodigies like Rubinstein, studies of brain-damaged adults, reports on idiot savants, cross-cultural accounts, as well as the child development literature, supports the inclusion of musical aptitude on our list of intelligences. Even though it runs counter to our first intuitions of what constitutes "intelligent" behavior, musical aptitude belongs on our list along with linguistic and logical-mathematical aptitude.

In the view of Multiple Intelligences, all seven faculties are equivalent -- some are not more "important" than others. Although twentieth-century western society values the linguistic and logical skills most highly and offers rewards to those who excel in these areas, other cultures value the intelligences differently. We must be careful to distinguish the psychological level, on which the intelligences are equivalent, from the sociological level, on which the intelligences may be differentiated.

Bodily-kinesthetic Intelligence

Movement of various parts of the body is controlled by the movement cortex regions of the brain, a localized function that is well-documented in the research literature. This control is contra-lateral: the right hemisphere of the brain is responsible for control of movements on the left side of the body and vice versa. Support for the claim that bodily-kinesthetic activities constitute an intelligence is supported by the fact that impairment of voluntary movements through conditions of brain damage can occur while reflexive movements of those same body parts can occur on a non-voluntary basis.

The bodily-kinesthetic intelligence is responsible for such activities as athletics, crafts, and dance. Although the intelligences are independent and distinct, in a task of any complexity, several intelligences are usually deployed in concert. For example, playing the violin, a task that taps the musical intelligence, also requires a sophisticated form of bodily-kinesthetic ability.

Interpersonal Intelligence

Interpersonal intelligence builds on the core ability to notice distinctions among others, in particular contrasts in their intentions, temperaments, moods, and motivations. This skill appears in a highly sophisticated form in religious and political leaders, teachers, and therapists.

The relationship between Anne Sullivan and Helen Keller illustrates the fact that interpersonal intelligence does not depend on language. Anne Sullivan, the "miracle worker," was herself legally blind and she was not trained in special education. Nevertheless, she successfully faced the daunting challenge of educating a blind and deaf seven-year old, an education that was further complicated by the emotional struggle the child was engaged in as she tried to understand the world around her.

The experiences of Anne Sullivan and Helen Keller underscores the interpersonal understanding that is a necessary part of all teaching. Also, this situation again reminds us of the difference between an intelligence, a cognitive capacity of the brain, and the modes of receiving information, usually the eyes and ears. For Helen Keller the visual and auditory modes were blocked, but she was able to obtain that information through the mode of touch. Although Helen Keller was impaired in some ways, certainly there was nothing wrong with her intellectual capabilities.

Intrapersonal Intelligence

This final capacity is responsible for understanding one's own internal aspects -- access to one's feeling life, range of emotions, as well as the capacity to discriminate among these and eventually to label and draw upon them as a means for guiding one's behavior. This intelligence is most private and can only be seen at work when expressed through one of the other intelligences, such as language or music.

At the age of 21, Langston Hughes dropped out of Columbia University and went to sea. The first night out, he threw all of his books into the ocean. One book fell into the scupper -- he climbed down, picked it up and threw it overboard with the others. Why? In his autobiography, Hughes reveals his motivations:

> It was like throwing a million bricks out of my heart -- for it wasn't only the books that I wanted to throw away but everything unpleasant and miserable out of my past: the memory of my father, the poverty and uncertainty of my mother's life, the stupidities of color-prejudice, black in a white world, the fear of not finding a job, the bewilderment of no one to talk to about things that trouble you, the feeling of always being controlled by others. All those things I wanted to throw away. To be free of. To escape from. I wanted to be a man on my own, control my own life, and go my own way. I was twenty one. So I threw the books into the sea. (Hughes, 1986, c 1940, p. 99)

This anecdote reveals the intrapersonal intelligence, the individual's self-awareness, as well as the personal courage in creating an unflinching expression of that understanding.

Implications for Education

The theory of Multiple Intelligences has a number of significant implications for education. In this section I will examine two of them: the importance of establishing a rich, meaningful context for problem solving; and the relationship between self-esteem and the full identification an individual's intellectual profile.

Context in Problem Solving

The theory of Multiple Intelligences reminds us of the importance of a "hands-on" educational process. In the arts and in the crafts, students learn by doing. To learn to paint, students paint; to learn to operate a table saw, they operate a table saw. In the humanities and in the sciences, in contrast, students learn almost entirely by reading and talking, rather than by doing for themselves. In history class, students read summaries of the work of historians; they don't "do" history. In English class, they read interpretations of novels and analyses of plays; they don't write novels or perform plays. In science class, students review the procedures and findings of pivotal experiments, they don't design and conduct their own experiments.

The theory of Multiple Intelligences suggests that there are a number of shortcomings when education is restricted as it is in the humanities and sciences. The heavily verbal context favors students who excel in the linguistic intelligence while at the same time it does not challenge students to pursue problems using the other intelligences. The exercises, problem sets, and examinations in school are all solved in the same, "school-like" way.

Because the problem-solving context in school is uniquely structured and largely linguistic, students often fail to transfer the problem-solving skills they are developing in school to situations outside school. On the job, for instance, a person is expected to solve a problem using any intelligence that yields a useful solution. By focusing on structured, linguistic solutions to problems, schools do not give students sufficient opportunity to develop the necessarily flexibility in thinking. In this restricted context, schools establish a special context for problem solving that does not reflect problem solving in the world outside school.

Self-esteem

Working in this restricted context, students often create a false sense of themselves. Some students, those who are most successful in school because of their linguistic facility, may find themselves with less of an advantage after they leave school. They have come to think of themselves as efficient problem-solvers, and yet when they encounter problems in an unrestricted environment, they struggle to find adequate solutions. Other students, often those who are less successful in school, find that they have very important skills for solving problems in the working world that went unrecognized in school.

Two hypothetical examples illustrate this disparity. First, think of a student who answers correctly and quickly on all school tests, regardless of subject area. This student is also a class leader and involved in many extracurricular activities. However, success in school for this student does not lead to similar success later. Indeed, it is not difficult to imagine this student in a working situation in which he fails to respond with facility, especially when the setting is highly ambiguous and the tasks have no "right answer." The student struggles in this setting, despite his success in school.

Next let's imagine a very different student, someone who is rather ordinary in solving school tasks, but who has the special skill of quick adaptation to new situations. This student also has superior interpersonal and intrapersonal intelligences and can efficiently mobilize these capacities in the world outside of school. Forming teams of workers, managing limited resources, and handling ambiguous tasks all come naturally to her. The second student surprises her high school teachers by her success in the work place.

This is not to suggest that no "A" student will succeed in the working world or that no "C" student will struggle. What we find in looking at large numbers of students, however, is that there is surprisingly little correlation between school success and success on the job. The "C" student is just as likely to be successful outside of school as the "A" student. The problem is that in rewarding one type of student and not the other, school raises the self-esteem of the favored group and lowers the self-esteem of the group that it does not favor. School tends to ignore the importance of certain intelligences, and in so doing it discriminates among students.

Multiple intelligences suggests that school need not be structured in this way. For instance, school can help students exercise their interpersonal intelligence by establishing settings for cooperative problem solving. In fact, research indicates that students working in groups actually learn more than individuals working alone. By recognizing students with superior interpersonal, spatial or bodily-kinesthetic skills, school can elevate the self-esteem of those students and provides them with a greater likelihood that they will apply those skills appropriately when they leave school.

Assessment from the Point of View of Multiple Intelligences

The theory of Multiple Intelligences instructs us to look carefully at the context of an activity as we try to understand individual proclivities. This need for rich contexts in problem solving extends to the task of assessing student learning as well. For example, if we want to evaluate an individual's skill in music, we ask that individual to play a piece on a musical instrument. If we want to assess a student's talent as a leader, we might observe that student interacting with her peers. From the performances that result from these situations, we can draw conclusions about what those students have learned about the art form or the social setting and we can generate some ideas about the specific intelligences that have been brought to bear.

In this section, I will examine the assumptions of traditional tests from the perspective of Multiple Intelligences; then I will outline an alternative called performance tests; finally, I will discuss the use of portfolios of student work with a focus on student reflection.

Traditional Tests from the Perspective of Multiple Intelligences

On objective tests, students read a question and identify the correct answer from a list of possible answers. These tests ask the student to exhibit a skill or reveal knowledge in the context of the test, not in the context of solving a problem in the domain. These tests rely heavily on sophisticated linguistic aptitude and performance on them can be seriously reduced for students who do not have this prerequisite linguistic skill. The results are usually reported in terms of the rank of the student within the population taking the test, not in terms of number of questions answered correctly.

These traditional tests alter the relationship of the student and the teacher to assessment. Since they rely on an external measure of competence or skill, these tests become the authority; neither the student nor the teacher has any responsibility for making a judgment of competence. In fact, both student and teacher are discouraged, even disallowed, from making this judgment. Because the results are reported as rankings, students compete; they do not demonstrate competence.

Consequently, tests do two things. First, they establish a very limited context for solving problems, one in which there are no tools, no materials, no collaborators, and a limited amount of time. The context consists entirely of a series of questions followed by correct and incorrect answers. Second, these tests assume all responsibility for measuring the intellectual capabilities of the students taking the test.

One problem with this approach to assessment is that it is entirely unique to the school setting. Once students leave school, they may never again take a multiple-choice test. After they leave school, however, students must learn to do for themselves precisely what the tests have been doing for them previously. Students must figure out what they are learning (or failing to learn). They must draw these judgments from tasks that are heavily dependent on context, in which there are no "right answers." They must adapt their performance based on these judgments. Furthermore, they do not have tests (or teachers) to help them make these judgments.

The theory of Multiple Intelligences reminds us why these two issues of assessment -- context and responsibility for assessment -- are important. Context reveals the intelligences at work. Responsibility for assessment exercises the intrapersonal intelligence in a way that makes the students independent learners and successful problem solvers after they leave the very special environment of the school.

Performance Assessment as an Alternative to Tests

Building on this view of assessment derived from the theory of Multiple Intelligences, researchers, including those at Project Zero, are exploring assessment techniques that are built around authentic performances. In music, for example, a teacher evaluates a student's facility with a given piece by asking the student to perform that piece -- the performance itself is the "test." The assessment is "authentic" because performance on the test draws directly on the skills that the student is trying to master. The student practices the performance piece repeatedly, taking the "test" until she has mastered it.

The performances that are selected for assessment must reflect the actual skills and competencies that are valued in the field. For example, authentic skills in chemistry class might include designing an experiment around a question, gathering evidence, analyzing the resulting data, and reporting the results in a coherent and convincing manner. An authentic task in social studies might include conducting original research, reviewing relevant information in the library, and creating a video documentary that represents the results. In each case, students would practice these skills repeatedly until they have mastered them.

One example of a performance task in high school chemistry, developed by Dale Wolfgram and Compton Mahase for the Connecticut State Department of Education, poses this problem to students:

> You will be given two samples of soda; one regular soda containing sugar and one diet soda containing an artificial sweetener. Your task is to identify each sample as diet or regular. You must base this decision on the physical or chemical properties of the two different types of soda. As in any chemistry experiment, you are not allowed to taste any of the samples. Come up with a list of at least three possible ways to identify the samples and explain why you chose them.

Students start the task alone. Then they work in small groups for brainstorming and experimenting. Finally, students finish the task alone, answering a similar question concerning salt and fresh water.

As teachers evaluate student work on the Soda Task, they consider whether students can identify the appropriate properties of the liquids for the purposes of identification; can identify the information and steps needed to solve the problem; and can communicate those strategies through written means. (Baron, 1991)

Portfolio Assessment

Taking the notion of performance assessment one step further, the evaluation of these performances and their artifacts can be extended by collecting them in portfolios. As students work through a number of performances, they collect the results in a folder. Later, they select from these artifacts a specific collection that "tells the story" of what that they have learned and the skills that they have mastered. This collection, along with a description of what has been selected and why, comprises the portfolio.

The portfolio collection should not be restricted simply to the student's best work. It should also include drafts, outlines, and early attempts, since these are equally important to the task of demonstrating what the student has learned and the specific skills and concepts mastered. Also, as the student looks back over the folder of work, selecting pieces for the portfolio, these interim pieces are an important element that fill in the "biography" of the process that the student went through.

A number of important things can happen with this portfolio collection. First, the portfolio captures the student's work over the entire course of the year. As an assessment it reaches well beyond the "snapshot" examination that captures only the student's knowledge and capabilities at a specific moment. The portfolio can encourage students to take risks, to explore novel solutions to familiar problems, and to attempt more difficult strategies that may require longer periods of time. The portfolio can also reveal patterns in students' growth and learning.

Second, the portfolios can link the students' work in school to the culture that surrounds the school. For example, if students are working in the community, they can use their portfolios to connect those efforts with their school work. For example, a high school student who is doing volunteer work in a hospital might use her portfolio to make connections between that volunteer work and her biology course. Without the portfolio, the two experiences may be disconnected; but by looking for points of contact over the course of the year and by documenting those connections in her portfolio, the student can demonstrate her learning about biology in an applied setting that is meaningful to her personally.

Finally, the portfolios encourage students to take ownership for their work and to reflect on their progress. Rather than simply hurdling a series of obstacles, students become increasingly responsible for establishing personal goals and then for demonstrating that they have reached those goals through a collection of work. To bring about this sense of ownership, students must consistently work with their portfolios, reviewing the materials that they contain, making selections for inclusion or exclusion, and analyzing and discussing their choices. Students should also take every opportunity to share their portfolios with peers, parents, teachers, and other interested adults. In short, the process of reflection and sharing amplifies the central importance of each student's portfolio and the work it contains.

Reflections by Students on Their Work

Student reflection is a meaningful ingredient in a portfolio not only because it fosters a sense of ownership, but also because it is instructive at the same time. Far more important than the specific facts and skills that students learn in school are the insights they develop into the learning process itself. Students must learn how to teach themselves new skills and ideas, because once they leave school, they will no longer have the guidance of teachers and tests. Formal schooling can foster this ability by having students pay careful attention to their individual learning styles, by having them make important choices about their learning while they are in school, and by having them create portfolios that document those experiences.

Of course, the portfolio approach with its reflective component will not be effective immediately and automatically. Students must learn how to create portfolios and how to think about themselves as learners. The portfolio must become part of the educational experience of the classroom and part of the regular conversation between the teacher and the student as well as among the students themselves. When this happens, the focus of the classroom changes and the relative roles of the students and the teacher begin to change as well.

Summary

The move from the theory of Multiple Intelligences to performance assessments is straightforward. In order to analyze an intelligence, we must find problems that put it to work. We cannot learn about an individual's interpersonal intelligence or about his musical intelligence by asking him questions. We must pose for that person an interpersonal problem or a musical challenge. If we simply ask questions, we are evaluating the linguistic (and perhaps the logical-mathematical) intelligence instead.

Furthermore, if we want our schools to prepare students for the challenges they will face after they leave, we must constantly pose challenges in school that force them to invoke a variety of intelligences. These challenges should have different kinds of solutions, they should involve a variety of intelligences, they should encourage collaboration, and they should provide opportunities for reflection. In other words, to make our assessments more compatible with Multiple Intelligences, we must make them more authentic and more oriented toward performance.

At the same time, we want to foster the intrapersonal intelligence as well. To do so, we must pose problems and situations for students that evoke performances, and then encapsulate the resulting work in portfolios and help the students reflect on that work. If students leave school with plenty of practice self-consciously solving many types of problems, they will be better equipped to solve novel problems in the working world by drawing on a more complete understanding of themselves and their strengths and weaknesses.

Implications of the Theory of Multiple Intelligences for Multicultural Education

Finally, we turn to the implications of this theory of for multicultural education. I raise two questions in this regard. First, do different cultural or ethnic groups manifest different intellectual endowments? Second, what does our analysis of school from the standpoint of Multiple Intelligences suggest for the bilingual student?

The Question of Intellectual Endowment

The question of whether intellectual endowment varies from one ethnic group to another is a particularly difficult one because it leads quickly to issues of bias. For example, I am occasionally asked if particular ethnic groups are more skilled in certain intelligences than others. One group might be especially musical and kinesthetic; another group might have special spatial skills; still another excels in the verbal realm. This brings quickly to mind the racial and ethnic stereotypes of the African-American athlete, the Irish politician, and the Korean science fair winner. My answer can be simply stated: there is no evidence to support intellectual differentiations based on racial or ethnic origins.

There is, of course, important variation in intellectual competence among individuals, both in the computational ability of each intelligence and in the combination of intelligences in the intellectual profile. However, membership in a particular ethnic group does not predict any of this individual variation. In any classroom, students will reflect a variety of intellectual profiles -- some students will be especially verbal, some interpersonal, some spatial, and so on. This intellectual variety appears in all classrooms; it does not matter if the students are all from the same racial or ethnic group or if they represent different groups.

Although the individuals vary, the various racial and ethnic groups have the same innate intellectual endowment that they manifest in different ways. For example, given the same linguistic intelligence, some groups rely heavily on written language, others favor an oral tradition and still others communicate through linguistic codes.

The fact that schooling relies heavily on particular forms of linguistic communication and administers examinations that are heavily dependent on a particular form of linguistic skill puts students from a different linguistic heritage at a disadvantage. Furthermore, this singular approach to language can establish a disjunction between the culture of schooling and culture of the community. The theory of Multiple Intelligences reminds us that this disjunction, which may make school irrelevant and alienating to students from a different linguistic tradition, is a feature of cultures and not of intelligences (Banks, 1988, 1989).

A similar disjunction between the manifestation of the intelligence in school and its manifestation in the community can occur for each of the other intelligences as well. For instance, studies in school tap the spatial intelligence in geometry and geography; the culture of the community, on the other hand, may value graphic design or chess playing. School places little value on interpersonal skills, while the community may value those skills highly.

In sum, there are important differences in how students from different cultural groups deploy the various intelligences and how the intelligences are valued by those cultural groups. One strategy for coping with these differences might be for school to reduce the distinctions between the use of intelligences in school and in the community; a second strategy is for school to find ways of demonstrating a respect for those differences and celebrating the individual competencies in students even when those competencies are different from the basic expectations of school.

Implications for Bilingual Education

As for the bilingual student, it should be clear by now that the highly linguistic environment of school, with its focus on written language, places at a disadvantage any student with difficulties in the linguistic realm. The ability to learn and the ability to display that learning are both impaired in the bilingual student in this highly verbal setting.

Perhaps the most important implication for bilingual education from the theory of Multiple Intelligences is the importance of separating the intellectual capacity from the skill with using the language of the dominant culture. Just as school often fails to recognize the abilities of students who are successful in the world after leaving school, it also fails to recognize the abilities of students who have not mastered the language of school.

One remedy for this situation is to provide more situations in which students can display competencies that do not rely as heavily on specific linguistic skills. Projects, in both the arts and in the crafts, can be an excellent indicator of these capabilities. Working cooperatively in groups is a second. Display of diligence or creativity over a period of time is a third. If we can build this variety into the school setting, we can more accurately identify students with talents and students with difficulties, apart from their mastery of language. We can make our schools more reflective of and better preparation for the world outside school. And we can give our students a more complete sense of themselves.

In summary, if we are to take Multiple Intelligences (and multiple cultures) seriously, then school must establish a meaningful context for problem solving; it must provide an opportunity for students to practice using a variety of intelligences; it must build self-esteem by helping students develop an accurate and complete picture of their capabilities; and it must establish assessment situations that facilitate and reinforce these ideas.

Schools that Provide Opportunities for Success

To a large extent school is a mechanism for transmitting the expectations of society and for sorting the members of that society. Because that transmission is based on language, the sorting is also based on language. The theory of Multiple Intelligences predicts that such an environment will place many individuals at a disadvantage and will unfortunately yield the view that not every student can learn. Indeed, with its focus on linguistic skill of a particular sort, traditional schooling consistently underestimates the capabilities of many very talented bilingual students. Indeed, this misrepresentation occurs for any student whose particular blend of intelligences does not match precisely what the traditional school requires.

There is an alternative. We might begin to think of school as a place where students pursue the successful accomplishment of meaningful activities rather than the locus of sorting and the gatekeeper to future opportunities. Schools for success must provide a variety of opportunities for students by considering the different intellectual proclivities and cultural predispositions that students bring to school.

Such a view takes seriously the notion that every student can learn; but it does not require that all students learn in the same way. Just as the musician and the backgammon player solve different problems and use different intelligences, they can both be remarkably successful at what they were doing but in very different ways.

Introducing multiplicity to this analysis and emphasizing success does not imply that school must lower its standards, that "anything goes." Quite the opposite is the case. Successful accomplishment requires genuine challenge, high standards, and definitions of accomplishment that are acknowledged publicly. Furthermore, we can bring demanding techniques of evaluation to these disparate activities via the assessment alternatives of performances, projects and portfolios. Using these techniques, the schools for success can document and evaluate a variety of performances while maintaining very high standards.

A school that evaluates on a normal curve is not a school in which all students can be successful, because only half of its students can be above average. In contrast, a school that respects and responds to the multiplicity of aptitudes, that builds on its students' bilingual backgrounds, and that allows for variety in student performance, can strive for success for all.

References

Banks, J., & Banks, C. (eds.). (1989). Multicultural education. Boston: Allyn and Bacon.

Banks, J. (1988). Ethnicity, class, cognitive and motivational styles: Research and teaching implications. Journal of Negro Education, 57, 452-466.

Baron, J. (1991, July). Performance assessment in mathematics and science: Scoring performance assessments. Paper presented at the Institute on Diversified Approaches to Assessment, Harvard University, Cambridge, MA.

Gardner, H. (1983). Frames of mind. New York: Basic Books.

Hughes, L. (1986). The big sea: an autobiography. London: Pluto.

Keller, E. (1982). A feeling for the organism. San Francisco: Freeman.

Rubinstein, A. (1978). My young years. New York: Alfred A. Knopf.

Soldo, J. (1982). Jovial juvenilia: T. S. Eliot's first magazine. Biography, 5, 25-37.

Response to Joseph Walter's Presentation

Vera John-Steiner
University of New Mexico

Although discussants are supposed to be either overtly or covertly critical, I have the pleasure of being enthusiastic instead. The two presentations that preceded my discussion have given powerful insights into the nature of knowledge acquisition and knowledge transformation. I plan to approach this issue with a similar spirit but with a slightly different data base and perspective.

When I first moved from New York to New Mexico, I was strongly committed to the central role of language in human thinking. This assumption reflected my European cultural upbringing where arguing and participating intensely in exchanging ideas around the dinner table seemed to be the most exciting thing a young child was allowed to do while joining his or her elders.

In contrast, I observed that Navajo and Pueblo children conveyed knowledge by dramatic play, by drawing, by re-enacting their experiences in spatial and kinesthetic ways. This observation was a challenge to my theoretical stance. It meant that I had to make a serious shift in my own approach to the nature of thought and theories of thinking. My approach is constructed within A Vygotskian framework, but a modified one, as developed in my book, Notebooks of the Mind. The impact of external activities -- such as computing -- upon the way in which we represent knowledge is central. In a culture where linguistic varieties of intelligence are dominant in the sharing of knowledge and information, verbal intelligence is likely to be widespread. In cultural contexts where visual symbols predominate, internal representations of knowledge will reflect visual symbols and tools.

One may think of schooling as a repertoire of resources, of culturally developed means to amplify one's own knowledge and intelligence. But if schooling only amplifies a limited set of knowledge representation (in our culture, verbal and mathematical approaches), learners are thereby restricted in using their own forms of intelligence, as the previous speaker described. My conception of intelligence is: those means by which we represent and transform received knowledge and prepare to contribute new knowledge.

Each of us is a subset of the total human possibilities. To develop our intellectual resources, we must focus our energies upon areas where we are most likely to be recognized as contributors, whether in our family, our preschool, or our communities. In studying creative individuals, I was impressed by how they chose to focus their attention on developing some of their strengths.

While our resources for education have been shrinking and while our stature as educators may have been diminished, our ideational fluency, our ability to come up with powerful new ideas has not been diminished. Indeed, it is now being nourished in new ways, partly because of our stronger commitment to cultural pluralism and to what I refer to as cognitive pluralism. My interpretation of the multiplicities of ways in which we represent knowledge does not have the strong biological base that Howard Gardner's theory of multiple intelligences does. Our approaches have in common our emphasis upon the diversity of knowledge acquisition and representation.

I would like to mention an additional point that has not been mentioned thus far. It concerns ways we create new knowledge in this last decade of the twentieth century. In the early decades of this century, Nobel laureates usually received the Nobel Prize for individual achievements. Ten or fifteen percent of them received a prize for collaborative work. Today well over two-thirds of the Nobel Prizes are given for collaborative work. Similarly, if you look at National Science Foundation applications, in the early years, most applications were by individual investigators. Now, 75 percent of all applications are either written collaboratively or include plans for collaborative execution of the project.

If we recognize that new knowledge is being developed through collaboration today far more extensively than heretofore, we must recognize the absolute necessity of learning how to work with complementary skills in group endeavors. And then we must recognize that the value of individualistic attributes such as IQ measures and other competitive assessments is rated out of proportion to its real social significance.

Currently we need to identify ways in which complementary intelligences are needed for joint endeavors that will contribute to the rapid development of new knowledge. By working from theoretical perspectives that emphasize multiple intelligences and cognitive pluralism, we must begin to pay serious attention to teaching and learning through projects, through cooperative learning, through interactional means. We are moving away from the traditional expectation that children on their own will master learning how to learn. To be a contributing member of our society means not only to assimilate knowledge but to communicate it, to share it, and in the process of sharing it, develop it further. The approaches to multiple intelligences that I find particularly exciting are revealed through ways in which individuals learn something about their own strengths and weaknesses through interaction with others. If you are asked to assess yourself in terms of Gardner's seven intelligences, you can reach only a first approximation of your talents. To go beyond this first approximation, you really need to test your hypotheses about your own abilities through interaction with others. The multiple intelligence perspective implies a much stronger emphasis upon assessment of authentic performances than do the measurements of individual IQs upon which formal academic gatekeeping has relied for so long. In authentic performance, you address a real audience and accept the constraints of a real environment. You not only demonstrate your own learning but also invite the consequences -- in terms of impact upon audience and the fit with the environment -- of that which you have learned.

Performance assessment also means monitoring your own growth over time. For such achievement, I find the portfolio movement very promising. It encourages the growth of that deep self-knowledge which I have found characteristic of individuals who have been successful in constructing a creative life. They worked out their own rhythms of productivity and of absorbed receptivity by developing a critical awareness of the conditions of their performances. The intrapersonal level of intelligence, or condition of introspection, is crucial because it provides information about your own rhythms of productivity, your own ways of determining when you need to work with others and when you need to focus on your own development. We can provide opportunities for such engagement from kindergarten on.

Self-knowledge depends upon interaction with others, a seeming paradox that we rediscover whenever we study the values of cooperation, collaboration, and communication of individual achievements to a receptive audience.

I think that the growing recognition that intelligence cannot be measured by a single distributional measure urges us toward recognition of cultural pluralism. Cultural pluralism provides us, particularly in this country, with the opportunities to really examine the implications of various ways of representing, transforming, and adding to knowledge. If we move in this direction, we are also providing opportunities for children from homes where English is not the primary language. We are encouraging these children to introduce into their school experience ways of knowing already characteristic of their home experiences where they frequently already share and represent knowledge across generations. We are encouraging engagement with diversity. We make this engagement not simply a glimpse into the alien world of an esoteric culture but active participation through which learners utilize their diverse ways of knowing to contribute through new communication designs.

I must tell you again how stimulating it is to be a discussant on this panel. Often, I have felt the discouraging effects of our need as educators to justify our existence through simplistic, even demeaning methods of assessment. Such effects are causing us to lose more and more potential educators, some the most promising members of our profession. We need to challenge such individuals, not discourage them.

Emphasizing cooperation and collaboration across as well as within generations highlights processes by which we can approach the extraordinarily demanding task of keeping the citizens of our society adequately informed rather than drowned in its flood of information. To achieve such coherent social engagement, we need to design curricula that are project based, that are vertically organized, that are cooperatively envisioned, that are linked to community concerns which extend beyond the confines of school walls. We need to utilize children's museums, local theater and other performance groups, community service agencies, a variety of available community resources -- utilize the many ways of learning that free us from the passive, frightened sitting that at present characterizes so much of our formal education.

Response to Joseph Walter's Presentation

Sue Teele
University of California, Riverside

My whole life is devoted to public education. I am married to a superintendent of schools, am a board member for the Redlands Unified School District, and an administrator at the University of California, Riverside. I work, live, and love public education. My goal is to make education the very best place for students and to enable all students to reach their fullest potential. Based on that premise, I would like to take you on a roller coaster ride, in twenty minutes, through what Dr. Walters has been saying about the Theory of Multiple Intelligences. I have been involved with the Theory of Multiple Intelligences for the last two years, and I believe, truly, that we have found a way to reach all students. As I look at education right now, I see having a window of opportunity for the next three to five years to make effective changes in public education. I would like to see the Theory of Multiple Intelligences and new methods of assessment be right up at the top of the list.

One of the reasons why we must change public education is because our students are very diverse with multi-faceted problems. I am going to share with you some statistics stated by California State Superintendent of Public Instruction Bill Honig concerning an average group of high school sophomores. In a class of thirty sophomores, Honig states that four will speak no English, eight are two or more levels below in math and reading, one is a victim of child abuse, three will be teen parents, three will grow up in public housing, eight will be on public assistance, seven will not graduate, and seven will not be employable. Now, those of you in the audience, who are logical-mathematical, can quickly add that up, and what do you find? It's more than thirty. What does that imply? We have some students that fit more than one statistic and have multi-faceted problems. What that means is we have to look strongly at what we are doing in public education, and ask the question, are we providing a quality education for all students? I suggest a change in philosophy. The change in philosophy of education is this: we must create an educational system in which an individual learning plan enables all learners to proceed at a rate and a pace that is challenging and achievable, makes no unfair comparisons with the progress of others, assures positive reinforcement and creates positive self-esteem for all students. This is not a simple sentence, and it is not an easy task. However, it is what I would like to see happen in public education, and I truly believe that the Theory of Multiple Intelligences is a way to do this. It is a way to reach all students.

Now, for those of your who are visual-spatial learners, I am going to show you some information about the seven intelligences. These are the seven, if you didn't have them memorized, take a look at them visually, so you can become familiar with them: Linguistic, logical-mathematical, intrapersonal, spatial, musical, bodily-kinesthetic, and interpersonal. I am going to give you a very quick run though all seven intelligences to help you have an understanding of them. If you want to informally assess yourself, please do so on a scale of 1 to 5. Everyone in this audience has all seven intelligences, so zero is not an acceptable answer. You have all seven, but each one of us is unique, because our strengths are in different combinations of the intelligences. We are a microcosm of every single classroom in our nation. That is why we must recognize the diversity of our student population in our schools and teach to that diversity. We must help all our students find ways to succeed. This is a critical component in order for effective change to occur in our schools.

Let me describe for you linguistic intelligence. If you are strong in linguistic intelligence, you have highly developed auditory skills. You like to read and write. You like to listen. Your vocabulary is well developed. You enjoy writing stories and using a computer for word processing and editing. You often spell words accurately and easily.

The next intelligence is logical-mathematical. If you are logically-mathematically intelligent, you explore patterns, categories, and relationships. You enjoy mathematics. You like to work with computers, not the word processing necessarily, but the problem solving, data base, and spread sheet aspects. You are able to group and order data and make interpretations and predictions. You prefer order in your life. You reason things out logically and enjoy problem solving to find solutions.

If you are intrapersonally intelligent, you have a deep awareness of your inner feelings, strengths, and weaknesses. You have strong opinions when controversial topics are being discussed. You prefer your own private inner world, and often, when given a choice, like to be alone rather than be with groups.

If you are spatially intelligent, you think in images and pictures. You like to draw, paint, and participate in art activities. You are able to report clear, visual images when thinking about something. Often, you can read maps, charts, and diagrams. You respond positively to movies, slides, pictures, and anything that has a visual image. Spatially intelligent individuals respond positively to a visual medium.

If you are musically intelligent, you are sensitive to a variety of sounds in the environment. Some of you were more sensitive to singing "Happy Birthday" than others. Some of you found yourself humming the song. Often, you like to have music on when you are studying or when you are working. We are conducting research at UCR in the area of musical intelligence and looking at what kinds of music are appropriate in classrooms. We are finding that Baroque music is pleasant for students. It is very relaxing and has a tempo the same rate as the heart beat. I did an experiment at the junior high level and let the students have headphones to listen to their own music while taking a test. Guess what I found? They couldn't concentrate on taking their tests because their music was distracting. More research needs to be done regarding how and when to create a musical environment in our schools.

If you are bodily-kinesthetic, this conference is difficult for you as you have to sit for long periods of time. It has been estimated that about 80 percent of our high school dropouts and between 60 percent to 80 percent of students in special education have bodily-kinesthetic intelligence as their most dominant intelligence.

Please understand that I am in charge of two special education programs and trained as a special education teacher, as I am going to make a statement that may upset some of you. I believe that many of our students who are in special education are not learning handicapped, that we in education are simply handicapped in teaching them how to learn. Because many students are dominant in bodily kinesthetic intelligence, they require active learning activities. Bodily kinesthetic individuals learn through their bodily sensations. They like to touch, feel, and tap things. They have difficulty sitting still long periods of time and thrive on hands-on active learning activities. They need manipulatives, role playing, simulations, physical exercises, competitive sports and action-packed stories. They require movement; and to sit in a classroom at a high school level five to six hours a day is very difficult for them; that may be why many drop out. We need to include many activity-based experiences in all our classrooms to engage more actively the bodily-kinesthetic students in the learning process.

The seventh intelligence is interpersonal intelligence. These individuals enjoy being around people. They have many friends and socialize everywhere. They enjoy participating in cooperative learning groups. Roger Johnson is a good friend of mine. We have discussed interpersonal intelligence and the relationship between multiple intelligences and cooperative learning. Interpersonal intelligent individuals have a lot of empathy for the feelings of others and can respond to the moods and temperament of other individuals.

Let me show you something interesting. As I said, I have worked with about 2,000 educators. I asked them to do an individual assessment of themselves and select their three most dominant intelligences. This information has been analyzed and we discovered that of the 2,000 educators, 17 percent were linguistic, 12 percent logical-mathematical, 19 percent intrapersonal, 10 percent spatial, 14 percent musical, 11 percent bodily-kinesthetic, 17 percent interpersonal. The great thing about this discovery is that not all educators are linguistic and logical-mathematical. As Dr. Walters was saying, we must teach to all seven intelligences in the classroom. Well, guess what! There's only a 9 percent differential between high and low with educators, and that is so exciting to me, because what that says is we can incorporate multiple intelligences into the classroom because we, educators, represent all seven intelligences in a diverse way. What we must do is represent all seven intelligences in every single classroom in this nation and that means our methodology must be very different. We must have a repertoire of strategies that we use.

Do you want to see something interesting? I have been observing elementary classrooms in Southern California. I designed a pictorial multiple intelligences inventory that is appropriate for elementary schools. I asked 600 kindergarten through sixth grade students to circle the picture that they thought was most like them. Let me show you what I discovered. I have been working at an elementary school that has a demographic profile of 76 percent Hispanic, 10 percent Black, 11 percent Anglo, and 3 percent Other. It's an interesting school to study as one-third of the students are LEP students.

I would like to discuss a graph that depicts the intelligences profile of kindergarten, first, second, and third graders. For some of you who may be in the back, these bar graphs represent linguistic, logical-mathematical, intrapersonal, spatial, musical, bodily-kinesthetic, and interpersonal intelligences. Here's what I found. At the kindergarten level, the number one dominant intelligence when I assessed the students on the pictorial inventory was intrapersonal intelligence. Number two was linguistic intelligence and number three was bodily-kinesthetic intelligence.

When I moved into the first grade, the findings indicated the number one intelligence was spatial intelligence. Number two was logical-mathematical intelligence. Number three was linguistic intelligence. In the second grade, I found the most dominant intelligence was bodily-kinesthetic, second was spatial, and the third was logical-mathematical. In the third grade, spatial intelligence was first, logical-mathematical intelligence was second, and bodily-kinesthetic intelligence was third. You will notice something very interesting. The kindergarten level was totally different from any other grade level. Let me show you something interesting. When I examined the fourth, fifth, and sixth grades, I discovered a very interesting pattern here. Let me remind you that first and third grades were most dominant in spatial intelligence, second grade in bodily-kinesthetic intelligence, and kindergarten in intrapersonal. When I studied the fourth grade, I found bodily-kinesthetic intelligence first, logical mathematical intelligence second, and spatial intelligence third. I studied the fifth grade and found spatial intelligence first, logical-mathematical intelligence second, and bodily-kinesthetic intelligence third. There was a direct correlation between third and fifth grades. Isn't that interesting? As I studied sixth graders, I found bodily-kinesthetic intelligence first, logical-mathematical intelligence second, and spatial intelligence third. These were the same three dominant intelligences as for fourth graders. The interesting thing is that logical-mathematical intelligence was second in grades 3, 4, 5, and 6.

After completing the inventory, I went back to the teachers and asked them to validate this information. They thought that these findings were very accurate and agreed that is where they perceived their students.

Those of you who are studying research will say, that's only one school studied and is only preliminary findings at one school. I agree. It is only one school, and I am going to work with several other schools this year to compare them with this one. I want to see if there are commonalities of intelligences between certain grade levels. I don't know what I will find. Stay tuned as it will be interesting to see. If I do find commonalities that are specific to certain grades, there may be important curriculum implications for elementary schools.

I recently began studying the middle school and found in a preliminary study with seventy 8th and 9th grade students that 6 percent were dominant in linguistic intelligence, 4 percent in logical-mathematical intelligence, 7 percent in intrapersonal intelligence, 12 percent spatial, 23 percent musical, 30 percent bodily-kinesthetic and 18 percent interpersonal intelligence. How do we teach at the middle school level? We teach using linguistic and logical-mathematical intelligence. What should we be doing at the middle school level? We should be teaching methodologies that reach all seven intelligences and, according to these findings, emphasize bodily-kinesthetic, musical, interpersonal and spatial intelligence. We must tap into the seven intelligences in order to get all students engaged in the active learning process.

I also recently studied a high school speech class and an ESL class. I also found spatial, musical, bodily-kinesthetic, and interpersonal intelligence as the highest intelligence in both classes. The only difference between the two classes was logical-mathematical intelligence. In the ESL class, the students, who were predominantly from Mexico, scored that intelligence as their second highest.

Education in America is at a turning point where it is important to accept the theory of multiple intelligences and incorporate into the instructional process the philosophy that all students can succeed. The concept of authentic assessment should be included as we have to look at assessing students very differently. We have had a lot of discussion on authentic assessment today. Portfolio assessment, performance assessment, scientific investigations, open-ended questions that allow the students to think and problem solve, and untimed and integrated testing are all a part of authentic assessment. Research is being conducted at Chico State University in California in regard to untimed tests in mathematics and its relationship to gender. They found that, when time was not a factor, there was no significant differences between boys and girls on their mathematics tests. Girls tend to respond to mathematics linguistically. Boys tend to respond spatially and logical-mathematically. As a result, girls take longer to solve problems in mathematics. How do we solve that? Simply allow students more time on tests. Encourage them to solve problems through their dominant intelligences.

Student self-assessment is absolutely essential. We have to involve students in their assessment process. In student self-assessment, students evaluate their own progress. They evaluate solutions to problems. They decide if they have contributed appropriately and made progress in their development and become aware of what they know. Are they aware of what they now know and still feel they need to learn? In authentic assessment, students must become actively involved in the assessment process. It is also extremely important that we combine instruction with assessment.

We are currently working at UCR on a project with three high schools in regard to college admission. These high schools will submit to the University, with their applications, portfolios from grades 10 and 11 for English and mathematics. We feel we may be able to learn more about their content level than we can learn only through a SAT score. We are going to track those students' progress for four years through UCR.

I would like you to remember this. Some of you may have seen this -- WYTIWYG (What you test is what you get). You get what you assess. You do not get what you do not assess. We need to be sure we build assessments that measure what we as educators teach and want taught. This is so important.

"The ultimate purpose of all strategies is to foster student achievement and engage students in the active learning process. We should emphasize individual differences in all their qualitative richness. This means that education should always provide for differences of interests, not just once in a while, but always, and not merely permit, but encourage diversity in the way students spend their time in school. It means to give students significant choice, to let them become responsible in every possible way, the regulation of their own learning," said Hawkins. William Glasser said something that I think is very important, "We learn 10 percent of what is read, 20 percent of what we hear, 30 percent of what we see, 50 percent of what we both see and hear, 70 percent of what is discussed with others, 80 percent of what is experienced personally, 95 percent of what we teach to someone else." Why can't we provide opportunities for students to teach to one another, if that retention rate is that high?

My work with the Renaissance Project has provided opportunity for me to observe some interesting things happening in the classroom. One of the most exciting observations was made in a study in a bilingual first grade classroom. In that classroom, I saw a non-English speaking student move very quickly into the world of reading and writing in English. Do you know why? Because the teacher discovered he was spatially intelligent and asked him simply to do spatially intelligent activities when he first entered. She didn't say, "You must read and write in English, right this minute." She said, "We are so happy you are here. Welcome to our classroom. Aren't you a wonderful artist!" What we found was that because the student's self-esteem was enhanced by the teacher and elevated by all the students in the classroom, he learned how to read and write in English because he was in a comfortable environment conducive to learning. That is what we must do in education.

In closing, I have a vision that takes us beyond the 1900s and into the year 2000. My vision is that we will change the philosophy of education so that all students will have an opportunity to reach their fullest potential. To do that, we have to change assessment and we have to move into the theory of multiple intelligences. We have to believe that every student is gifted and can succeed. We have to provide staff development to everyone involved. We need to be involved in legislation because, I believe, we in education should be telling legislators what needs to happen in education. We must recreate the thirst for education with all students, and we must provide opportunities for students to succeed. That is our absolute responsibility as educators. It is true that children are 25 percent of our population, but they are 100 percent of our future. We in education can make the difference in children's lives.

(Editor's note: Dr. Teele has developed a teachers' Certificate in the Study of Multiple Intelligences program at Riverside.)

Improving Bilingual Education Programs Through Evaluation

Alan L. Ginsburg
U.S. Department of Education

Introduction

This paper identifies how evaluation and research can contribute to federal policies for language-minority students who have limited proficiency in English. Federal bilingual education policies are controversial, and the debate benefits from systematically collected information about, and impartial evaluation of, bilingual programs.

Many of the concerns about bilingual education are misguided. For example, some people fear that bilingual policies will produce the divisive separatism that characterizes some multilingual nations, but research demonstrates little foundation for such fears in the United States. As Pease-Alvarez and Hakuta (1992) point out, by the third generation, immigrants, regardless of their nationality, have assimilated into American society, losing the ability to speak their home language fluently.

A related concern, also without foundation, is that bilingual education programs propose to substitute the home language for English. But the issue of whether immigrant children should learn English is not in question. All sides agree that English- language proficiency is essential for survival and gainful employment in America.

Concern about what instructional methods are best for achieving English-language proficiency, however, is valid. Historically, education was important to new immigrants, but not essential for success. Heavy industry and manufacturing provided plentiful well-paying jobs for people willing to do the hard work. Immigrant populations made their way upward through such employment. For America in the 1990s, however, relatively few high wage, unskilled jobs exist, and their number is continually shrinking.

The amount of education that is required to obtain a good job is increasing; a high school diploma is no longer adequate. During the 1980s real earnings declined for workers with no more than a high school education. In 1973, among men 25 to 34 years old, high school graduates earned 87 percent of what college graduates earned but, in 1987, they earned only 67 percent as much as college graduates.

Some postsecondary education has now become almost as essential for well-paying jobs as a high school diploma was 20 years ago.

[illegible] educational performance of American students relative to those of other industrialized nations. Current initiatives would replace these fragmented efforts with systemic reforms built around national education standards and national examinations. These changes would move the United States closer to the apparently more successful educational systems of our economic competitors and, if experience is any guide, would probably benefit language-minority and other students at risk for failure in school.

Education is essential to the economic success of language minorities, but the successful education of these people and other nontraditional populations is critical for our nation's economic well-being, too. Ethnic and racial minorities will account for about 30 percent of new labor force entrants over the next decade. Moreover, as the United States seeks to compete with other nations, the ability to understand and speak other languages becomes a resource to be developed.

This paper examines issues of evaluation and assessment in language-minority education within this broader context of education and its influence on the nation's future. The discussion is divided into three parts. The first part examines what has been learned from the evaluations of bilingual education conducted by the federal government during the 1980s. The second part assesses the implications of national standards and examinations for language-minority [illegible]

[illegible]

[illegible]

The Supreme Court ruling disallowed "submersion," a policy that [illegible]

The Lau remedies proposed by the federal government at the close of the Carter administration sought to further clarify school district responsibilities to LEP children. Under this proposal, school systems were to assess the relative proficiency of language-minority students in English and their native language. Instruction, at least in elementary schools, would have to be provided through a student's stronger language. Although the Reagan administration withdrew the proposed regulations shortly after entering office, the deep-seated divisions over the proposed rules pointed clearly to the need for studies to evaluate systematically and rigorously the merits of alternative approaches to language instruction.

After the withdrawal of the Lau remedies, the national debate shifted to Title VII of the Elementary and Secondary Education Act. This legislation specifically aimed to make students proficient in the English language. But the legislation also recognized the importance of instruction in the native or dominant language "to the extent necessary to allow students to achieve competence in the English language."

To help inform the debate, the Department of Education's Office of Planning, Budget, and Evaluation [illegible]

[illegible]

Longitudinal Study of Bilingual Education

[illegible] evaluation of three approaches to helping students who speak a language other than English (Ramirez, Yuen, Ramey & Pasta, 1991). The three approaches represented different degrees of exposure to English-language instruction, each reflecting a different philosophy for helping LEP students move into English-language classrooms.

In English-language immersion programs, the teacher uses English for all instruction while using the home language informally, as for occasional clarification or directions. The teacher obviously needs a working understanding of the home language but may not be fluently bilingual. Students may use the home language in responding to the teacher or talking to each other. Pupils are mainstreamed into English classrooms as soon as they have shown adequate proficiency in English.

"Late-exit" transitional programs are designed to help students become proficient in their home language before they develop proficiency in English. The teacher is fluent in both languages. Children entering elementary school receive several years of instruction in the home language. At about the fourth grade the instruction shifts gradually toward English. Students are not mainstreamed into the regular English classroom until grade 5 or 6.

The "early-exit" program is a transitional bilingual education [illegible]

[illegible]

Significant findings include the following:

- [illegible] demonstrated greater- [illegible] though language-minor- [illegible] to progress more slowly [illegible] enabled the students to [illegible] classrooms. Nonetheless, [illegible] remained considerably below the norm for other students.

The pattern of English-language progress for students in late-exit programs differed from the others in predictable ways. Late-exit students were initially less proficient in English. By fourth grade, about half of the students exposed to English immersion and early exit instruction were rated by their teachers as good or very good in English language comprehension, compared with [illegible] percent of late-exit students. By sixth grade, [illegible] percent of the late-exit students were so rated. (Comparable sixth-grade data were not collected for immersion and early exit programs because those students typically no longer received special language instruction.) Of some importance was the fact that the rate of growth for students in late-exit programs was increasing, although there is no way to project this trend to assess whether these students would actually approach grade-level norms.

- Teachers used ineffective methods of language instruction. Regardless of the method of language instruction, students had few classroom opportunities to produce language. Teachers did most of the talking in class. When students did interact with teachers, half the time they produced no language [illegible] (e.g., they were listening [illegible] k, they typically an- [illegible]

- [illegible] programs strongly sup- [illegible] their preference for [illegible] y associated with [illegible] Spanish. More than 90 [illegible] program wanted their [illegible] English. With respect to [illegible] parents of children in [illegible] permitting Spanish to be [illegible] half of parents of children [illegible] the parents of children [illegible] favored a particular in- [illegible] preference for instruction [illegible] language preference was [illegible] language instruction [illegible]

Virtually all parents (about 90 percent or more, regardless of the type of program) want bilingual teachers. This finding may reflect the parents' preference for teachers who are able to understand their children and themselves.

- Parental involvement is facilitated by instruction in the home language. More parents of children in late-exit programs monitor their children's homework (74 percent) than do parents of children in immersion or early-exit programs (53 percent). Parents may be more comfortable with teachers or better able to help their children when instruction is given primarily in the home language.

- Students typically come from environments in which both Spanish and English are spoken; this circumstance may explain why mixed-language approaches are effective. Parents of LEP children speak to each other in Spanish 86 percent of the time and to their children in Spanish 79 percent of the time. However, their children speak to their brothers and sisters mostly in Spanish only about 40 percent of the time. More homes receive English-language newspapers than Spanish-language papers (e.g., 45 to 37 percent), children spend 84 percent of their TV-viewing time watching English-language programs and 66 percent of their record-listening time listening to English-language records. Students also come from communities in which their neighbors are as likely to use English as Spanish.

These findings suggest that focusing evaluations on determining a single best method of language instruction for non-English-speaking children was probably the wrong approach to take to evaluation. Most special language programs in the United States represent a blend of different approaches. Indeed, the study had difficulty locating either late-exit or immersion programs, and the seven immersion programs in the study were all that could be found in the entire country. The fact that all three approaches could be effective for elementary school children indicates that the most important requirement is to learn one language well. That language does not initially have to be English, so long as transition to English occurs by the third or fourth grade.

Nonetheless, the fact that students failed to catch up to expected norms suggests that other factors, including program content, need greater consideration. Exposing language-minority and other children at risk to a more challenging curriculum is one goal of advocates for stronger national academic standards.

Bilingual Education and the Movement Toward National Standards

During the 1980s the policy debate over the appropriate method for instructing LEP students shaped the evaluation process. Little attention was given to the content of what was being taught. In the 1990s, however, evaluations of programs for language-minority students will be shaped by the outcome of the policy debate over whether this country should adopt national education standards. Proposals such as those in the Education Department's AMERICA 2000 initiative call for systemic reforms; these include setting national standards that establish what students are expected to know in core subject areas.

The final report of the National Council on Education Standards and Testing (NCEST, 1991), a congressionally created body drawing bipartisan representation from Congress, the administration, governors, teachers unions, and education experts, helped move the nation toward national standards:

> In the absence of well-defined and demanding standards, education in the United States has gravitated toward *de facto* national minimum expectations, with curricula focusing on low-level reading and arithmetic skills and on small amounts of factual material in other content areas. Most current assessment methods reinforce the emphasis on these low-level skills and processing bits of information rather than on problem solving and critical thinking. The adoption of world-class standards would force the Nation to confront today's educational performance expectations that are simply too low.

The report's conclusions are consistent with the views of most Americans. Surveys demonstrate strong public support for accountability and national tests: 7 percent favor a standardized national test, 68 percent a standardized national curriculum, and 81 percent national goals and standards.

With broad public support and evidence from other industrialized nations on the effectiveness of standards, the United States is likely to move toward some system of national standards and examinations soon. The implications of these changes for language minority students need to be carefully explored. Concerns about the fairness of tests for language-minority and other at-risk populations could be magnified under a high-stakes national examination process.

The experience with minimum competency testing indicates that standards need not have harmful effects. When these tests were instituted during the mid-1970s, there was some concern that the requirements would hold minority students back and cause more of them to drop out of high school. But trends in student performance indicate that competency standards probably worked to the benefit of students from nontraditional backgrounds.

The National Assessment of Educational Progress represents one of the best sources of consistent information on student performance since the 1970s. In 1975, only 52 percent of Hispanic 17-year- olds read at the basic proficiency level; in 1988, 73 percent did. And the proportion who read at the adept level in 1988 (24 percent) was nearly double the proportion who read at that level in 1975 (13 percent). In addition, between the mid 1970s and 1990, Hispanics' scores on the Scholastic Aptitude Test (SAT) improved by 28 points, while white students' scores declined by 9 points. Although Hispanic dropout rates remain unacceptably high, they appear to have declined slightly since the mid 1970s.

Despite these gains the performance of Hispanic students remains below the level for white students, and the gap worsens at higher skill levels. Because competency requirements seemed to have previously benefitted at-risk students, raising requirements through new national standards and encouragement could further extend these benefits.

To work, however, national standards must be perceived as fair, must seek to challenge and motivate students to improve, and must provide students with the special resources needed to improve. Ultimately, the success of a system of national standards will depend on answers to the following questions:

- When is it appropriate to test children from non-English language backgrounds? Children exposed to English for the first time presumably need a transition period before testing. Conversely, students must not be excluded from testing for so long that schools are no longer held accountable for their performance.

- In what language is it appropriate or even feasible to administer the test? Issues of feasibility, accuracy, and appropriateness have to be resolved. How feasible is it to translate tests into languages other than English, and what is the cost of doing so? Can a student's stronger language be accurately determined? Is it appropriate to test knowledge of the English language while testing knowledge of the content of other subjects in a student's stronger language?

- How can test results be used to expand student opportunities rather than simply to punish students who are experiencing difficulties? Testing can reinforce students' educational opportunities, if schools use test results to identify and correct student weaknesses. The NCEST has proposed testing students in grades 4 and 8 as well as grade 12 in order to detect and correct deficiencies. Such a test pattern would differ from the practice of most other industrialized nations, which test students only once before tracking them into college programs.

Early identification of problems, of course, does not guarantee that needy students will receive special support. Schools should be required to address special problems as a condition of testing. Moreover, if schools are failing, they should be held accountable. Many schools, particularly ones serving lower income areas, are insulated from pressures to provide high-quality education to all children. Recent legislation included in Chapter 1 of the Elementary and Secondary Education Act requires schools that fail to meet performance goals to institute a performance improvement plan. A Chapter 1 type of improvement plan could be extended to cover schools failing language-minority students.

- How can the tested material be coordinated with a challenging curriculum? A valid criticism of current standardized testing is that the material on which students are tested may never be taught in school. This circumstance puts at-risk students at a particular disadvantage, because these students are least likely to be exposed to the range of general-knowledge questions on standardized tests. Aligning course content and tests with curriculum frameworks would give at-risk students a fairer chance.

New standards would have implications for federal evaluation requirements under Title VII. The current Title VII legislation requires local programs to report an almost impossible amount of information: subject areas taught; instructional methods; time spent on specific tasks; preparation, language abilities, and educational background of the staff; students' achievements in English language arts and subject areas, oral proficiency in English, and achievement in native language; each school's grade retention rate, dropout rate, absenteeism, number of referrals to special education, number of placements in gifted and talented programs, and postsecondary education attendance.

Faced with excessive reporting burdens, recipients of federal bilingual education grants have simply ignored most of them. A 1990 evaluation independently assessed the quality of Title VII evaluation reports. Although most programs used appropriate achievement tests, fewer programs analyzed test data appropriately. Only about half used a 12-month testing interval, although use of shorter test intervals is known to seriously overstate gains in student achievement. Less than a quarter of the programs reported test data in sufficient detail to draw programmatic conclusions. Finally, very few programs (about 15 percent) followed former participants to assess their progress in the regular education program, although this assessment may represent the best measure of program effectiveness.

Instead of being a paper exercise, local evaluations of federal bilingual education programs should become an integral part of program operations. Evaluations should focus on the performance of students in relation to national standards, and the quality of local program evaluations must improve considerably.

Implications for the Federal Role

As already mentioned, two sets of issues have been explored in the evaluation of programs for language minorities: in the 1980s, the focus was on instructional processes, while in the 1990s the focus is on instructional content. These two evaluation streams need to be combined in a coherent strategy that integrates the content of what is taught and the methods of instruction.

The upcoming re-authorization of Title VII offers an opportunity to debate and craft legislative responses that build on evaluation evidence and new educational reforms. Although the details of reform will require careful analysis, here are five general principles that could help guide reforms:

1. Bilingual programs should be held accountable for high achievement by their students, while local programs should be allowed flexibility over the method of bilingual education. Evaluations have demonstrated that bilingual education can work, but that no one method is uniformly superior. Successful programs may focus on dual language development or may immerse children in English immediately. In return for strong accountability for student performance, the federal government should expand local program discretion over federal resources. For instance, federal legislation discourages programs from serving students for more than three years. If student performance is satisfactory, there is no reason to limit the length of bilingual education programs.

2. Teachers of LEP students in bilingual and regular classrooms need sound training. Evaluations have shown that even teachers in thoughtfully designed programs appear to use pedagogies that are not effective. The federal government's Title VII program should become a major source of teacher training support, but this support should ensure that the training provided is sound and likely to take hold in a school. Bilingual education training, now focused almost entirely on teachers in the bilingual program, might be extended school-wide. Because all teachers in

the school work with language-minority children, all could benefit from training in language instructional approaches.

3. Accountability requirements should shift from traditional standardized tests to performance-based examinations that promote opportunities for language-minority and other at- risk populations to achieve "World Class" standards. Current standardized tests are not well coordinated with the curricula or services. Teachers perceive these tests as having little value and as being primarily punitive. A system of national standards tied to examinations must be linked to curricula. Poorly performing students should receive special help to enable them to reach the standards. Furthermore, schools that consistently fail such students need to be held accountable for this failure and not permitted to continue to operate on a business-as-usual basis.

Language-minority children should be excluded from testing only if they enter school with limited English proficiency, and then only for a specific period. Widespread exclusion would serve to stigmatize excluded students and diminish schools' accountability to provide the students with appropriate educational services.

4. The federal government should launch a multiyear agenda to identify best practices within different instructional approaches, rather than attempting to determine a single best approach. The evaluations of bilingual education in the 1980s sought a single winner to the question of identifying effective methods of language instruction. This approach was wrong. Evaluations for the 1990s need to be driven by the question of what approach works best under what conditions.

Research should also focus on strategies to encourage students to learn English outside school and to foster parental involvement in their children's education. These efforts should build on evaluation findings that show that language-minority parents will become more involved in education when schools communicate with them in their home language.

5. Federal bilingual education policy should recognize that the home language is a resource to be developed. Achieving bilingualism through foreign-language instruction for native-born Americans is an accepted national priority, one that is becoming more important in an increasingly competitive economic environment. Logically it follows that students who want to maintain their home language should have the opportunity to do so. A knowledge of the home language is not a substitute for strong knowledge of English but a recognition that knowledge of the home language and of English can help the development of both languages.

References

Baker, K., & de Kanter, A. (1981). Effectiveness of bilingual education: A review of the literature. Washington, DC: U.S. Department of Education, Office of Planning, Budget and Evaluation.

Cziko, G. (1992). The Evaluation of Bilingual Education: From Necessity and Probability to Possibility. Educational Researcher, 21(2), 10-15.

Mullis, I., Owen, E., & Phillips, G. (1990). Accelerating academic achievement. Princeton, NJ: Educational Testing Service.

National Council on Education Standards and Testing (1991). Raising standards for American Education. Washington, DC: Author. (Submitted to Congress, the Secretary of Education, the National Education Goals Panel, and the American People)

Pease-Alvarez, L., & Hakuta, K. (1992). Enriching our views of bilingualism and bilingual education. Educational Researcher, 21(2), 4-6.

Ramirez, J., Yuen, S., Ramey, D., & Pasta, D. (1990). Final report: Longitudinal study of immersion strategy, early-exit and late-exit transition bilingual education programs for language-minority children. (U.S. Department of Education, contract no. 800-879-0156). San Mateo, CA: Aguirre International.

Language Testing Research: Lessons Applied to LEP Students and Programs

John W. Oller, Jr.
University of New Mexico

> ...man is not just a creature of accident, chained to and formed by the particular cave in which he is born....No real teacher can doubt that his task is to assist his pupil to fulfill human nature against all the deforming forces of convention and prejudice... Moreover there is no real teacher who in practice does not believe in the existence of the soul, or in a magic that acts on it through speech (Allan Bloom, 1987, *The closing of the American mind: How higher education has failed democracy and impoverished the souls of today's students*, p. 20).

For educators at large, probably the first and most important lesson learned from language testing research is that *language proficiency (whether it is construed as a general factor or as a constellation of related abilities) is important in one way or another to nearly everything that takes place in education -- whether at school or elsewhere.* Language proficiency is a critical element in the process of becoming literate and all of the other public manifestations of human intelligence that enable us to be the social beings that we are. It is important to intrapersonal and interpersonal performances of all sorts. Language, perhaps more than any other aspect of our existence, is what enables us to be members of a community that includes people other than ourselves. Perhaps I can be forgiven, as someone who comes partly from a foreign language teaching background, for stressing as enthusiastically as I do that proficiency in another language is like a key that opens a door to new worlds of understanding and provides access to new communities. However, if we remain in a permanent state of *monolingual myopia* -- which in its most pernicious form is a terminal disease -- language can be a wall that separates us from all the world beyond our particular primary language community. To the *terminally monolingual*, the wall is invisible, intangible, and seemingly non-existent. Yet is it as impenetrable as solid granite and forms a prison more secure than concrete and steel ever could. Electronic surveillance in the prison is altogether unnecessary because the inmates are as unaware of their situation as Plato's inhabitants of the cave were of theirs.

The good news, of course, is that by acquiring a language or two beyond our primary linguistic system, we can become more aware of our limitations, prejudices, and the inevitable ignorance that plagues all the denizens of all the caves, and to some extent, we can, it seems, escape the special prison of monolingual prejudice. With this desirable aim in mind, the insight that I want to develop—that *language proficiency is central to all aspects of education* -- if it can be called an insight, will be news to no one in the bilingual education arena. Nor is it apt to make headlines with teachers who work with students of limited English proficiency (LEPs). Still, it is an insight that bears scrutiny and certainly criticism, and it epitomizes, I believe, what language testing research has to offer to a conference on evaluation and measurement issues relative to LEP students and the programs that aim to serve them. With respect to the evaluation of programs, a special sort of assessment problem, I concur with Prestine (1990) where she cites Rist (1982) who notes that program evaluation inevitably entails a general question that "is at once disarmingly simple and incredibly complex" -- namely, "What's going on here?" (Rist, 1982, p. 440, and Prestine, 1990, p. 288). I'll try to show that language proficiency is a critical element in answering this general question not only in relation to individual students but also with respect to program evaluation.

For the particular group of educators assembled at such a conference as this one, I doubt it will be necessary to sell the idea that *language proficiency matters*. This is something that I assume we all agree on from the start. We may differ, however, in subtle and unanticipated ways on just how language proficiency matters and to what degree it matters. What I will attempt to do, therefore, is to elaborate on the ways in which language proficiency seems to matter according to the evidences afforded by theory and research. My analysis will be based on a selective review of the relevant literature. Underlying all of the discussion will be the ultimate aim of reaching some practical conclusions concerning how we ought to go about testing and evaluation in educational programs for LEP students. The best I can hope for is to affirm some of the good things that are already happening, to offer some constructive (I hope) criticisms concerning theories and practices that need mending, and to encourage us to capitalize still more on the rich linguistic resources that are coming to us in ever greater quantities from a pluralistic world of many languages.

To that end, I would like to suggest that the first corollary of my starting premise, that *language proficiency is a central element of all educational undertakings*, might be that the term "limited-English-proficiency" implies a complement of "almost-unlimited-proficiency-in-some-other-language-or-languages." While I do not want to deny the benefits (or importance) of students acquiring a high degree of proficiency in English in these United States, I do want to suggest that it is strange that our educational systems and national policies (as diverse and amorphous as they may be; see Prestine, 1990, for a discussion of great interest) seem generally determined (at least in practice) to either ignore or to deliberately remove rather than to nurture and preserve the linguistic resources that are literally walking into our schools at an ever increasing rate. Corresponding to the common emphasis on limitations, disabilities, disorders, disablements, disenfranchisements, etc., it seems to me that there ought to be greater consideration of the positive complements of these terms. In this suggestion, I concur with Lynda Miller (1990) where she contrasts her emphasis on "competencies" (taking her cue from the term "multiple intelligences" as employed by Gardner, 1983 and seq.) with the more common "approach in which the emphasis is on deficits and disabilities" (p. 2) or on "impairments, handicaps, and disorders" (p. 4).

According to the positive complement of the "deficit approaches" -- which might be properly called "empowerment approaches" -- the attainment of language proficiency is perhaps the main road to social empowerment (Cummins, 1986). As Miller puts it (following Hirsch, 1987): "being literate...is possessing shared background knowledge and holding positions of responsibility and power at the macro-levels of society" (Miller, 1990, p. 3). David Olson (1986) goes so far as to suggest that intelligence itself is hardly more than "literate competence" (p. 338) or "the distinctive forms of symbolic systems evolved and exploited by a culture as a means for representing and acting on the world" (p. 345).[1] Even Walters and Gardner (1985) who think in terms of "multiple intelligences"; also see Gardner, 1983, 1989, 1990) say that in their later development "children demonstrate their abilities in the various intelligences through their grasp of various symbol systems" (p. 15). In fact, each separate intelligence, of the seven they advocate (which we review in part 3, below), is eventually seen "through a *symbol system*: language is encountered through sentences and stories, music through songs, spatial understanding through drawings, bodily-kinesthetic through gesture or dance, and so on" (p. 15).

These ideas, though not identical with the view that I would like to advocate and develop here, still point, as I understand them, in the direction we ought to follow, and all of them tend to show the central importance of symbolic systems of which, I will endeavor to show (following C.S. Pierce [1839-1914]), natural language systems are chief. At any rate, all of the foregoing provides, I hope, a suitable preamble, a jumping off place, for the development of my main argument which follows in four parts which I will preview immediately.

I begin with (1) a review of the history of primary and non-primary language testing and with a provocative question: how come there is no field of primary language testing? This quandary, will be resolved early in the discussion in a way that illustrates my starting point above about monolingual myopia. It turns out that there are in fact many approaches to the measurement and testing of primary language skills, but that nearly all of them have been mis-identified as pertaining primarily to some other actually incidental purpose. This was unlikely to be noticed, however, owing to the pervasive monolingual myopia that has been prevalent for more than a century of public schooling and that still pervades the American educational scene. Until research on the testing of non-primary language proficiency began to bud in the late 1950s, hardly anyone ever thought to ask about research into the character of primary language proficiency. For this reason, the ideas to be gleaned from non-primary language testing especially, may be of some use to educators at large as well as those who work with the growing numbers of LEPs in our schools.

In order to see the connections of research in non-primary language measurement with broader issues in education, the second major section of this paper is a review of (2) the broader literature of educational measurement as it relates to the central theme -- the critical role of language proficiency. We will view that theme from a variety of angles and try to develop an up-to-date idea of where we are at present with respect to the unwieldy problem of measuring LEP students and evaluating the programs that purport to serve them.

The third major section of the paper offers (3) a somewhat elaborated idea of the nature of language proficiency [illegible] gence and distinct multiple intelligences. I will argue that these theories are not incompatible, but rather that they are complementary ways of viewing different kinds of distinctive human abilities.

Finally, I will conclude with (4) a few observations about how we might go about the practical business of testing (and also of teaching) the increasing numbers of LEP students that are working their way through our schools. I will recommend deep rather than surface assessment through discourse-based, real life performances.

(1) Research in Primary and Non-Primary Language Testing

In undertaking a review of research on language testing, as soon as we begin to talk about "non-primary language testing" we are bound to ask: Why is there no distinct field of primary language testing? The answer to this question is that many approaches to the business of measuring primary language skills do in fact exist, but that they go by many different names. For instance, "intelligence testing" generally aims at primary language proficiencies and "verbal intelligence testing" specifically does so. Measures of listening and speaking abilities, speech and hearing tests, literacy tests of all sorts, but especially tests of reading vocabulary, reading comprehension, and writing proficiency tests clearly aim at primary language skills. In addition to the traditional categories of intelligence and achievement tests, there are many deficit oriented categories of primary language assessment: e.g., tests of "language disorders," "learning disabilities," "mental retardation," and more recently many different sorts of "cognitive" and "metacognitive" tests, not to mention "linguistic" tests, "sociolinguistic elicitation devices," tests aimed at "discourse abilities," "grammatical intuitions," "metalinguistic awareness," etc. I submit that there are many reasons why these various approaches to primary language assessment have not been recognized as a coherent branch of educational measurement, but none, I suppose, is more important than the general affliction of American educators with what I am calling here, monolingual myopia. I hasten to add that I am not saying that there are no important differences among the various fields of study listed in this paragraph, nor am I suggesting that primary language proficiency is the only object of [illegible]

[illegible]

[illegible] and the school tests of "achievement," "aptitude," and "intelligence" [illegible]. The same complaint against traditional approaches to the study of intelligence in particular is what has led Gardner ([illegible]) and his collaborators (also see Walters and Gardner, [illegible]) to develop the theory of "multiple intelligences." However, I submit that if it was the prevalence of monolingualism among the American educators that held the reigns of power from the early part of this century that set them up to misinterpret a mere lack of proficiency in English as a second language as a widespread intelligence deficit among children and adults from non-English speaking backgrounds. As Hakuta (1986) shows, immigrants in the early decades of the twentieth century were often described as "linguistically confused," "mentally retarded," "learning disabled," and so forth. By now it is clear that measures of yet to be acquired language skills were simply misidentified as indicating deficient cognitive powers of a much deeper sort.

Moreover, as Ortiz and Yates (1983) have shown, the problem is far from solved as we approach the twenty-first century. In Texas alone, as recently as eight years ago, Ortiz and Yates found that Hispanics were grossly over-represented (about 300 percent) in classes for the mentally retarded and other exceptionalities. Interestingly, as Cummins (1984) points out, the American Association of Mental Deficiency still depends on IQ scores (formerly one but now two standard deviations below the mean) as a part of its definition of "mental retardation" (McKnight, 1982). But why should anyone expect Hispanic children to have a 300 percent higher incidence of mental retardation than other ethnic groups in Texas? What most of those Hispanic children obviously have in common is Spanish rather than English as their first language. A small percentage of them, probably no greater than the percentage in other ethnic groups, may have some form of genuine mental deficiency, but there is every reason to suppose that the vast majority of Hispanic children in Texas are quite normal in their general mental abilities. Because so many of them, however, have been misidentified as exceptional we may suppose that some children with genuine difficulties have also been overlooked and are not getting the special educational they need.

[illegible]

[illegible]

[illegible] according to [illegible] of Binet, the best way to form an impression of a child's intelligence was to "talk to him" ([illegible], p. [illegible]). In fact, the Binet and Simon ([illegible]) tests included such obvious language proficiency tasks as responding to commands (e.g., "Point to your nose"), repeating a phrase or sentence, naming objects, telling what's going on in a photograph, answering simple questions (e.g., "What's your name?" "Are you a boy or a girl?" etc.), counting coins, copying a phrase or sentence, reading aloud and recalling points of information, writing phrases from dictation, defining words, etc. All of this is relatively harmless so long as the language of the testing is the child's primary language system, but when it is not, difficulties arise. The nearly complete confounding of language proficiency with native intelligence persisted in the thinking of Binet who seemed to vacillate between the view that intelligence was distinct from acquired skills (Binet and Simon, [illegible], p. [illegible]) or that it was something that developed with "instruction" (p. [illegible]). In the year of his death he wrote that children of higher standing manifest their "intellectual superiority" mainly "in tests where language plays a part" (p. [illegible]).

The confounding of language proficiency with innate intelligence was especially apparent in a variety of fill-in-the-blank (cloze procedure) used by the German psychologist, Hermann Ebbinghaus [1850-1909]. According to David Harris (1985), as early as 1897, Ebbinghaus applied cloze procedure (more than half a century before its formal christening by Wilson Taylor, 1953) to meaningful prose with the intent of measuring the intelligence of school children. In the venerable tradition of *Gestalt* psychology, Ebbinghaus contended that intelligence involved linking elements so as to form coherent wholes. As paraphrased by Whipple (1915), Ebbinghaus is reported to have said:

> To measure intelligence, therefore, we must employ a test that demands ability to combine fragments of isolated sections into a meaningful whole. Such a test [that he called *Kombinationsmethode*] may be afforded by mutilated prose, i.e., by eliding letters, syllables, words, or even phrases, from a prose passage and requiring the examinee to restore the passage, if not to its exact original form, at least to a satisfactory equivalent of it (p. 285; as quoted in Harris, 1985, p. [illegible]).

[illegible] to Harris, claimed to [illegible] isolated sentences. [illegible] rather than connected [illegible] thus creating a near [illegible] scoring. While Trabue's [illegible], in my estimation, a [illegible] Trabue was among [illegible] measuring "language [illegible] Trabue-type fill-in tasks [illegible] afterward to be applied in [illegible] supposed to be measures, not of acquired language skills, but of innate abilities (e.g., tests by E. L. Thorndike, Lewis M. Terman, and others).

Subsequently the various tasks recommended by Binet and others were reinterpreted, and alternately amplified and reduced several times, and were eventually canonized into various modern IQ tests (Binet and Simon, 1905; Terman, 1916; Terman and Oden, 1947; Terman and Merrill, 1960; Kaufman, 1979). The best known examples of IQ tests are divisible roughly into the categories verbal and non-verbal (or performance) tests. In the non-verbal category *Raven's Progressive Matrices* and *Cattell's Culture Fair Test of Intelligence* are often used. Batteries aimed at both categories, however, are also well known (e.g., the *Thorndike-Lorge*, the *WISC-R*, the *Otis-Lennon Test of Mental Abilities*, etc.

Arthur Jensen of UC Berkeley fame (cf. Jensen, 1969, 1980) and Richard Herrnstein (1973; also Herrnstein and Wagner, 1981) of Harvard, extended the IQ testing movement, it would seem, to its most extreme limits by claiming to be able not only to reliably determine innate intellectual capacities but to distinguish races and ethnic groups according to such measures. Most thinking persons find their reasoning spurious and their claims unconscionable—a kind of intellectual atavism harking back to racist theories of the philosopher Nietzsche and the idea of an intellectual aristocracy promoted in relation to the eugenics movement that began with Sir Francis Galton (1869). While such views have been severely criticized (and, I believe, properly so; see Mercer, 1973, 1984; and Gould, 1981), the best argument against them has largely been overlooked: namely that what the traditional intelligence tests measure best are acquired primary language skills. This idea is latent in the recent literature on "multiple intelligences," but has rarely been brought to bear as some believe it should (Oller, 1991). For instance, Walters and Gardner (1985) say, "We speculate that the usual correlations among subtests of IQ tests come about because all of these tasks in fact measure the ability to respond rapidly to items of a logical-mathematical or linguistic sort" (pp. 13-14). This very nearly amounts to saying that what those tests mainly measure is primary language proficiency (Oller and Perkins, 1978).

In spite of the long history of primary language testing from the early 1900s forward under the guise of IQ measurement, the notion of language proficiency per se, would progress little until empirical studies of foreign language proficiency began to appear in the late 1950s. Among the first was Carroll, Carton, and Wilds (1959) showing that cloze procedure had some potential as measures of language proficiency. A spate of studies would soon follow (Carroll, 1961; Lado, 1961; Valette, 1964; 1967) but it would not be until that latter part of the 1960s that non-primary language testing research would begin to flourish (cf. Upshur, 1967; Upshur and Fata, 1968; Spolsky, 1968a, 1968b, Anderson, 1969; Upshur, 1969a, 1969b; Oller, 1970; Oller and Conrad, 1971; Savignon, 1971). From there forward, too many research reports, conferences, and books would be generated for them to be adequately covered in any single review. However, it would not be until June, 1984 that the first issue of the journal *Language Testing* would appear. By then certain general themes and trends had been fairly well defined and the many of the paths that are currently being followed out had been marked off. Rather than try to plod through the whole terrain, in what follows I will concentrate on what I think the most important themes were in the 1970s and 1980s and still are in the 1990s.

It was John Carroll (1961) who suggested the distinction between discrete point approaches and integrative approaches to language testing. Discrete-point tests were grounded in the taxonomic approaches to linguistics that would later fall into disfavor as the Chomskyan revolution (see Chomsky, 1956, 1957, 1965, 1972, 1975, 1980a, 1980b, 1988) began to have its fuller impact into the 1970s and 1980s (see Newmeyer, 1980). Discrete point tests were based on inventories (taxonomies) of various sorts of elements. For instance, the phonological system of a language was supposed to consist of phonemes which could be tested one by one. The lexicon was a list of words, and grammar (alias syntax) was a list of patterns. This taxonomic way of looking at language, and at human abilities in general, still prevails among many (though certainly not all) psychologists (cf. the numerous examples cited by Cummins, 1984), speech-language pathologists (Coles, 1978, and Cummins, 1986, document this claim), and educators in general (Cummins, 1984, 1986; Cummins and Swain, 1986; Bloom, 1976; Bloom and Krathwohl, 1977; Swanson, 1988).

According to the discrete-point model, a sufficient number of items aimed at elements drawn from the several inventories of phonemes, morphemes, lexical items, and syntactic patterns would assure a valid test of language proficiency. In the 1980s, this same taxonomical thinking would persist in lists of "notions" and "functions" of speech acts and discourse (cf. Farhady, 1983b, and his references). The latter extension was certainly a natural one, but it did not really depart from discrete-point theory. The purest varieties of such thinking, e.g., Lado (1961) contended that language test items should focus on only one skill (e.g., listening), and only one domain (e.g., phonology), and only one element (e.g., a particular phonemic contrast) at a time. Besides distinguishing domains of structure -- phonology, morphology, lexicon, and syntax (semantics and pragmatics were not much thought of during the discrete-point heyday) -- discrete-point testers also distinguished skills (listening, speaking, reading, and writing). It was claimed that a test item could not be very good if it mixed several skills and/or domains of structure. And this contention itself pointed to what Carroll (1961) called "integrative tests."

For instance, Robert Lado (1961) contended that giving dictation, a foreign language testing technique popular with language teachers (cf. Valette, 1964; Finocchiaro, 1964), was not a good method because it mixed everything together. It was integrative rather than discrete-point (i.e., taxonomical) in its orientation. According to Lado, dictation did not test phonemic contrasts since these were apt to be given away by lexical or syntactic context. It did not test words because the words were "given" by the person reciting the material to be written down. It did not test syntax since the syntax also was "given." Worse yet, according to discrete-point thinking, dictation mingled listening comprehension with writing and reading. It also mixed phonology, vocabulary, morphology, and syntax (not to mention semantics and pragmatics) into a potpourri.

Discrete-point theory, however, in the final analysis was more of a hypothetical perspective than a practical one. Had it been influenced much by empirical evidence, it would have had to be radically revised since language students in taking dictation do make many errors in just the domains that Lado claimed were not tested. For instance, in actual dictation protocols, we find evidence of phonemic contrasts that have been obliterated, for example, "collect" is apt to be rendered "correct" by an Asian writing a dictation in English. Or, complex consonant clusters of certain types of morphological inflections are apt to be omitted in many cases. Furthermore, the same persons who make these sorts of errors in taking dictation are apt to make analogous errors in writing an essay, speaking, or other discourse processing tasks. In fact, such problems carry over into relatively routine tasks such as repeating sequences of heard material, reading aloud, or even copying a text.

Also, in taking dictation, word order is sometimes adjusted in surprisingly creative and ungrammatical ways. Lexical items are changed radically. For example, in one study at UCLA a passage on "brain cells" was rendered in an almost coherent way by one non-native speaker of English as a text on "brand sales." Almost everything in the text was changed though a superficial phonetic resemblance remained between what had been dictated and what was written down. Less dramatic transformations of the same sort are commonly observed in dictation protocols (cf. Oller, 1979, pp. 283-285, for several examples).

As I argued in 1979 (p. 266) and continue to believe today, discrete-item tests do not accord well with what people do when they process text or discourse in normal ways. An example of a test exemplifying early discrete-point, taxonomical theory that has been widely applied but without much success is the Carroll and Sapon (1959a) *Modern Language Aptitude Test* (also see their *Manual*, 1959b). Carroll (1967) found, in a massive study of college foreign language majors near graduation, that the *MLAT* was only a significant predictor of foreign language attainment if extraneous variables such as interest, parental language background, and travel to the foreign country were included in the regression equations. Even with these extraneous variables added in, the *MLAT* still accounted for a modest 9 percent or less of the total variance in foreign language attainment. The several subtests of the *MLAT* itself, however, accounted for less than 1 percent of the total variance in foreign language attainment. More recently, Goodman, Freed and McManus (1990) again found the *MLAT* to be a non-significant predictor of success in foreign language courses for 586 students tested at the University of Pennsylvania. They speculated that perhaps the failure of the *MLAT* in this case was due to the fact that language teaching seems to be moving more and more in the direction of integrative, whole language approaches.

It is possible to find many examples of integrative tests that actually proved more robust both in theory and in practice than discrete-item tests. These included dictation (Valette, 1964), essays (Briere, 1966), answering questions orally (Upshur, 1967, 1969a), telling a story (Politzer, Hoover, and Brown, 1974), giving a speech, conversation or oral interview (ETS, 1970), reading aloud (Kolers, 1968), answering questions about a text (Politzer, Hoover, and Brown, 1974), repeating sequences from a text or narrative (also known as "elicited imitation"; Baratz, 1969; Politzer, Hoover, and Brown, 1974; Swain, Dumas, and Naiman, 1974), translating from L1 to L2 or the reverse ("elicited translation"; Swain, Dumas, and Naiman, 1974), etc. One of the various integrative types of task experimented with in the late 1960s and early 1970s was cloze procedure -- a method christened as such by Wilson Taylor (1953, 1956, 1957) for measuring readability of texts. It involves omitting words from a written (or possibly oral text) and requiring the examinee to replace the missing items (Anderson, 1969; Spolsky, 1968; Oller and Conrad, 1971; Oller, 1973).

As empirical research began to accumulate in the 1970s and into the 1980s it became clear that there were practical as well as theoretical differences between integrative and discrete-point tests. Integrative tests were apparently measuring some traits and abilities of language users that discrete-point tests could not get at. Still, even into the 1970s there were some, Earl Rand of UCLA, for instance, who insisted that discrete-point methods were either better or at worst equivalent to integrative tests (Rand, 1972, 1976). These claims were rarely sustained in practice. If one had examined closely the empirical results, it would have become clear that greater reliability and greater validity generally accrued to tests falling toward the integrative end of the spectrum.

Farhady (1983a) disagreed with this claim, but his examples were, as Oller (1983b, p. 321 footnote a) pointed out, drawn from tests that were quite integrative in character. Therefore, when Farhady (1983a) claimed that there was no difference between integrative and discrete-point tests with respect either to reliability or validity, he was really saying in effect that there is little difference between several about equally integrative tests. He was comparing reasonably good oranges with other reasonably good oranges. There were no truly discrete item tests in the inventory he compared. In any event, it is illogical to argue that the kind of test item that fully isolates a particular phonemic contrast, or a single lexical item, or a particular grammatical morpheme, or a syntactic rule, will yield results equivalent to the sort of test that requires the employment of a vast system of such relationships -- a whole grammar. If those two types of tests did turn out to be equivalent (which they are not, see also Damico and Oller, 1980; and Damico, Oller, and Storey, 1983), the result would be entirely anomalous as there simply is no theory whatever that predicts such an outcome. If a given phonemic contrast, say, /r/ versus /l/, is not in some sense distinct from, say, the syntactic transformation that copies the number of a referring head noun onto its respective present tense verb and its demonstrative modifier, e.g., in "These recommendations are...", then the distinction between phonology and syntax must be misguided. But how? While tests of particular phonemic contrasts, or inflectional morphemes, or syntactic rules, might generate reliabilities in the range of .6 to .7 (e.g., Evola, Mamer, and Lentz, 1980), tests of a more integrative character generally yield reliabilities about 10 points higher in the range of .8 to .9 (Oller, 1972, for instance). Or consider the fourteen different integrative tasks used in research to calibrate the language question on the 1980 U. S. Census, none yielded a reliability lower than .98 (cf. Scott, 1979).

It seemed to many, therefore, toward the end of the 1970s that integrative testing had prevailed over discrete-point approaches. However, this conclusion may have been premature. In the context of normal language processing, any given discrete-point item of interest may always be singled out for special attention *in that context*. On the other hand, a single element of any sort (a thoroughly isolated discrete-point) in the absence of the dynamic tensional context of discourse is like the sound of one hand clapping. Such discrete-points become mere fictions, like the dimensionless points of a line. Without the line, the points along it are dimensionless locations occupying space exactly nowhere. In context notions of discrete elements of language structure or skill are valuable theoretical constructs, but without the context, they are undefined fictions.

Out of the controversy over discrete-point versus integrative tests, there emerged a distinction of a different sort. While the original dichotomy (proposed by Carroll, 1961) was based on superficial aspects of test items, domains of structure, and modalities of processing, it became increasingly clear that the distinction had been incompletely and inadequately drawn. Carroll (1961), Rand (1976), and Farhady (1983a) all observed that there never was a truly categorical difference between discrete-point and integrative test items. The difference was merely one of degree. The dichotomy formed a continuum whose end-points were fully distinct only in theory. In practice, there are no completely discrete-point tests anymore than there are points or lines in the space-time continuum apart from some object or trajectory to define them. In actual experience all test items are more or less integrative in character.

Normal language use always involves meaning beyond the theoretically discrete elements of surface forms. That is, there is a linking with persons, places, things, events, relations, etc., in experience. However, if this meaning aspect beyond surface form is admitted, no test item can meet the demands of discrete-point theory. As I have hinted several times above, it may be worth saying straight out at this point that semantics and pragmatics were notably absent from discussions of discrete-point items. This was probably due to the fact that meaning as such is never a discrete-point affair. It cannot be since meaning spills over into the whole continuum of experience which the very existence of meaning both presupposes and implies.

Another insurmountable difficulty for discrete-point theory was that language use occurs in real time and is therefore time-constrained. This is not so obviously true for reading and writing as it is for listening and speaking tasks. However, it is easy to prove with a little thinking that in fact there are severe temporal constraints on reading and writing as well as on oral tasks. Meanings that involve long-range constraints in a written text, for instance, are essentially inaccessible to persons who lack a certain level of language proficiency owing to the limited time that they can hold the target language material in working memory. If the requisite part of the memory image fades from consciousness before the part with which it must be linked can be grasped, it will be impossible because of this temporal fact to grasp the full meaning.

Moreover, there are many other ways that real time constraints operate with reference to reading and writing in respects that are precisely analogous to temporal constraints on oral tasks. For instance, we may not have time to go and ask someone what So-and-So's last name is so we can look him up in the phone book. Or, we may not have time to drive to the library to look up a particular reference for a research paper. We may spend hours looking for a certain statement in a large book, or several volumes. These cases are hardly different from the problem of trying to recall some significant detail from a conversation (e.g., did he say to turn right or left on Oak Street?). In the final analysis, the salient differences between speech and writing seem less so when we look more closely at each one. Time and meaning, respectively, constituted the pragmatic naturalness constraints that led to a differentiation, therefore, of a certain subclass of integrative tests that came to be known as pragmatic (Oller, 1973, 1979; Cohen, 1980; Savignon, 1983). This subclass, it turned out, was entirely distinct from discrete-point tests. In fact, the pragmatic naturalness criteria eliminate any strictly discrete-point item as unnatural. Such items do not really involve normal language use anymore than the recitation of a number or parroting a numerical operation constitutes mathematical reasoning.

In addition, many tests that are thoroughly integrative in character also fail to meet the pragmatic naturalness criteria. For instance, the proofreading test explored by Barrett (1976) was integrative but failed the meaning criterion. It involved the omission of morphologically redundant elements (e.g., plural markers, tense indicators, articles, prepositions, verb particles, etc.) from prose and required the restoration of these elements by examinees. A peculiarity of the task was that fluent readers had to attend so much to the surface form of the text in order to notice the missing elements that they failed to process the meaning of the text and after performing the task could not even tell what the text was about. On the other hand, *examinees who did concentrate on the meaning, and who could answer reasonable questions about its content,* would invariably get low scores. These results are consistent with the frequent observation by proofreaders that plying their trade slows down their reading. In fact, they often resort to rather unusual methods of checking surface forms such as reading the text backwards, or following it word-for-word while someone else reads aloud, and the like. These extreme measures are useful because proofreading requires a somewhat unnatural attention to surface form and good readers are often the worst proofreaders because they supply much information that is not in fact in the surface forms at all (cf. Goodman, 1967; Goodman and Goodman, 1977; Goodman, Goodman, and Flores, 1979; Smith, 1975, 1978, 1982, 1984, 1989).

Another procedure that is integrative but fails the time requirement is the sort of multiple-choice cloze test where a list of many (say, 50 or more) words are given and must be reinserted, one by one, into a text with blanks. This task is highly integrative but may involve looking back and forth between the list and the text, and a constant rereading of the list. It may be more like solving a cross-word puzzle than normal discourse processing. Because of the frequent interruptions, in looking back and forth between text and list, and the time lapses while reading the list, it is doubtful that such a task constitutes a pragmatically viable procedure. At any rate, as the list of possible words becomes longer and longer, it is clear that the task resembles less and less the normal processing of discourse.

What was more important about pragmatic tests, and what is yet to be appreciated fully by theoreticians and practitioners is that all of the goals of discrete-point items, e.g., diagnosis, focus, isolation, etc., could be better achieved in the full rich context of one or more pragmatic tests. As a result, it was argued that the valid objectives of discrete-point theory could be completely incorporated within a pragmatic framework. However, the goal of separating each and every element of structure or skill from the whole fabric of experience was abandoned. As an analytic method of linguistic analysis, the discrete-point approach may have had some validity, but as a practical method for assessing language abilities, it was misguided, counterproductive, and logically impossible to achieve.

Another outcome of the discrete-point/integrative controversy, and the empirical research which it spawned, was a reconsideration of the almost forgotten *g*-factor of Charles Spearman (1904, 1927). This development had two sides: one statistical and the other theoretical. The statistical side of the argument was soon resolved against any all inclusive *g*-factor, but the theoretical argument has yet to be adequately considered.

Charles Spearman had observed that most intelligence tests, in his day (and it may be noted that things have changed little since then; cf. Jensen, 1969, 1980) were strongly correlated. By inventing factor analysis, then a new statistical technique, Spearman showed that it was possible to identify a single general factor underlying most IQ tests and accounting for a huge chunk of variance in all of them. The same argument could still be extended to almost all achievement, competency, and proficiency tests used in education today (see Oller and Perkins, 1978, Gunnarsson, 1978, and Stump, 1978, and for counterpoint and response, Carroll, 1983b, and Oller, 1983a; but see Gardner and Hatch, 1989 who claim to be able to measure separate "intelligences" independently). This general factor came to be known as "*g*" or "the *g*-factor". Subsequently, L.L. Thurstone (1924, 1938, 1947; also Thurstone and Thurstone, 1941) and others, argued in favor of a plurality of primary mental abilities instead of a single *g*-factor of intelligence. They never settled how many primary factors there were or just how to define them. They vacillated in the end between six and eight distinct primary factors. In more recent years Guilford's "structure of intellect" model has multiplied these factors to 120 (Guilford, 1967). More recently still, Gardner (1983, 1989, 1990), Gardner and Hatch (1989), and Walters and Gardner (1985, 1986a, 1986b) have picked up the cudgel again on behalf of multiple intelligences. While Gardner and colleagues differ in their particular list of "intelligences" from the "primary factors" proposed much earlier by the Thurstone's, there is a fundamental resemblance in both the arguments and applications of the ideas favoring profiles that look at the broad spectrum of a person's abilities rather than a single IQ score.

However, long before Howard Gardner and colleagues came to the fray, it was generally admitted (by L.L. Thurstone himself, and more recently by his student J.B. Carroll and others) that underlying any set of primary factors or secondary or tertiary ones there will still be a general factor. A recent study of language proficiency by Fouly, Bachman, and Cziko (1990) concludes that a second order general factor and a model that allows differentiated components at the first order level are both fairly good at predicting observed relations between different language measures for 334 ESL students at the University of Illinois. They refer to Carroll (1983a) who summed up both his results and those of Fouly, et al. (1990) in terms of the long term controversy over general versus specific factors in language testing research:

> With respect to whether the results support a "unitary language ability hypothesis" or a "divisible competence hypothesis," I have always assumed that the answer is somewhere in between. That is, I have assumed there is a "general language ability" but, at the same time, that language skills have some tendency to be developed and specialized to different degrees, or at different rates

> so that different language skills can be separately recognized and measured (p. 82).

Fouly, et al. go on to say, "the present study provided support for the differentiated skills hypothesis recurrent in the works of Bachman and Palmer (1983), Carroll (1983a), Farhady (1983c), and Upshur and Homburg (1983)....Similarly, the findings of this study support the claim that, in addition to differentiated language skills, there exists a general factor" (p. 16). In support of the latter model they might have cited Oller and Perkins (1978, 1980) and Oller (1983a). A general factor of language proficiency (or what has been called "intelligence," in the case of tests of primary language abilities), cannot be denied on statistical grounds (Carroll, 1983a, 1983b).

While at first multiple factors as contrasted with a general factor were thought of as mutually exclusive, this was never correct. The general factor, whimsically referred to as the Godzilla factor by Purcell (1983) could be useful in spite of the fact that it did not exhaust all of the reliable variance in a number of language tests and even though could be transformed in a variety of ways into a multitude of component factors (see Farhady, 1983c; Upshur and Homburg, 1983; Bachman and Palmer, 1983; Vollmer and Sang, 1983). Godzilla, therefore, was prematurely proclaimed to be dead (by Purcell, Farhady, and others), and certain persons set out to bury him (Alderson and Hughes, 1981; Palmer, Groot, and Trosper, 1981; Porter, 1983; Spolsky, 1983; Alderson, 1983; Hughes and Porter, 1983; Davies, 1984). But Godzilla refused to be buried. It was true that he was not quite tall and strong enough to embrace the whole world (i.e., explain all of the variance in all tests), but he was plenty large and strong enough to resist burial (Bachman and Palmer, 1983; Carroll, 1983a; Bachman, 1990; Fouly, Bachman, Cziko, 1990; Oltman, Stricker, and Barrows, 1990).

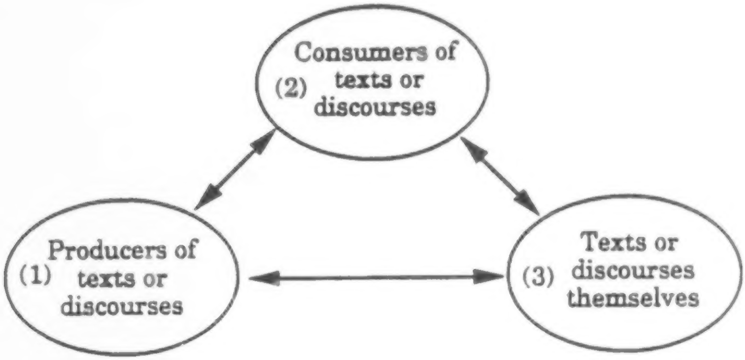
Although some researchers continue to pursue the elusive goal of resolving the general factor into its "proper" components (Sang, Schmitz, Vollmer, Baumert, and Roeder, 1986; Bachman and Clark, 1987; Bachman, 1990; Fouly, Bachman, Cziko, 1990), it would seem that a definitive division of language proficiency into its contributing components may be unachievable in principle by virtue of the fact that the multi-faceted semiotic hierarchy can be viewed from many complementary angles that logically should prove to be about equally correct (witness the findings of Fouly, et al. 1990). At any rate, the most important side of the argument is not statistical, but theoretical -- the fundamental problem is to find a coherent theory and it is certain that this cannot be achieved by purely statistical methods (see Bachman, 1990, pp. 296-358; Cummins, 1981; Krashen, 1981, 1982, 1985; Carroll, 1983a, 1983b; Upshur and Homburg, 1983). Upshur (1979), Carroll, and others have shown that the componential resolution of a general factor into a plurality of contributing components is not at all incompatible with the notion that language proficiency may be a fairly coherent and integrated totality. If we consider the meaning of total scores on tests with diverse subtests, or if we consider the fact that communicative abilities interact in complex ways to produce composite results, it is clear that both general and specific factors *must* be present in language proficiency. We will examine a few possibilities in section 3 below in this paper.

Aside from exploratory and confirmatory factoring of the traits (or theoretical constructs) that we may posit as aspects of human mental abilities or language skills (which I do not take to be the same thing, contrary to Boyle, 1987) and methods associated with particular tests, a number of interesting research reports using item response theory (IRT; following Rasch, 1980; see Davidson, 1988; Lynch, Davidson, and Henning, 1988; and Kunnan, 1990;) or multidimensional scaling (Oltman, Stricker, and Barrows, 1990; and Oltman and Stricker, 1990 following Guttman, 1965) have appeared. The common purpose of much of the research has been to sort out distinct sources of variance in language test scores. Among the widely recognized possibilities are three major sources as shown in Figure 1 below: (1) producers of discourse or text themselves differ in language abilities (and other mental abilities as well), as do (2) consumers, and as do (3) the texts or discourses (items in the case of many tests) that are both produced and understood. These three sources of variance can, of course, be further parsed up in a great variety of ways. One of the interesting and instructive avenues of research has been item response theory (IRT). Citing a single study will show how IRT can be applied to turn up unexpected sources of test item biases.

Kunnan (1990) demonstrated with an IRT approach (using a one parameter Rasch model with approximately 844 subjects) that subjects of different native language backgrounds and gender differ in performance on certain language test items depending in part on the instruction they have received probably in their major fields of study. At any rate, *differential item functioning* (DIF) was observed on the 150-item ESL Placement Examination at UCLA used in the Fall of 1987 on about 15 percent of the items. Apparently, Davidson (1988; see footnote 1 on p. 742 of Kunnan, 1990) had already shown that the test items in question met the requirement of unidimensionality in order for one parameter IRT to be applied. Based on that assumption, Kunnan found that certain grammar items focussing on the definite article, one or more prepositions, and verb tense were easier for Chinese and Japanese subjects (than for Spanish or Korean subjects), though different items (three in each case) performed differentially for the two groups. Also four vocabulary items proved significantly easier for Spanish speakers: *hypothetical*, *implication*, *elaborate*, and *alcoholics*.

Figure 1

The Three Main Sources of Variance in Language Test Scores


Consumers of texts or discourses
(2)
Producers of texts or discourses
(1)
Texts or discourses themselves
(3)

Since these words have Latin bases and cognates in Spanish with similar meanings, Kunnan credited native language background itself with the observed DIF for these items. Additional differences were observed for gender on 20 items some of which seemed to differ according to the major field of candidates. Items oriented toward the sciences seemed to favor males. Three items that favored females could not be accounted for. The results are interesting insofar as they show that items may be unintentionally biased against or in favor of certain groups. However, remedies for preventing this sort of bias are not clear: Kunnan, for instance, recommends that "a broad range of test content and formats" may help to reduce instructional bias. As for gender and native language biases, these are more difficult to deal with. They can be spotted on a post hoc basis with IRT, and the items can then be rewritten, but it is not entirely obvious how the author's recommendation that demographic data be elicited in advance might be used in test preparation. Certainly for items that remain unexplained even after the post hoc IRT, a demographic questionnaire or any sort of pre-screening even by members of the targeted examinees would seem unlikely to avoid the, for the moment, unexplained DIFs. The research is, in my view, nonetheless important as demonstrating the subtle kinds of test biases that can arise and the widely different sources variance that may constitute such biases.

Similar, though somewhat more specific biases for Japanese learners of English as a foreign language are demonstrated experimentally by Chihara, Sakurai, and Oller (1989). Our work used a more traditional repeated-measures approach but predicted in advance what sorts of items in a cloze passage were biased against Japanese learners of EFL. Because Japanese subjects were compared against themselves in a repeated measures design, the variance of interest in particular items can be attributed specifically to the cultural or experiential background of the subjects tested. Two cloze passages were each presented in two forms: each passage appeared in an unmodified (biased) form and in a modified (reduced bias form). The method of modification was to change unfamiliar place names in the U.S. and Greece to familiar ones in Japan, and one instance of a mother kissing her son was changed to hugging (which is acceptable in Japanese culture). The results showed a significant advantage overall favoring the modified texts in spite of the fact that all else was left unchanged. The results, though based on an entirely different experimental procedure, agree with those of Kunnan (1990) using IRT, in showing that items may function differentially according to the background of subjects.

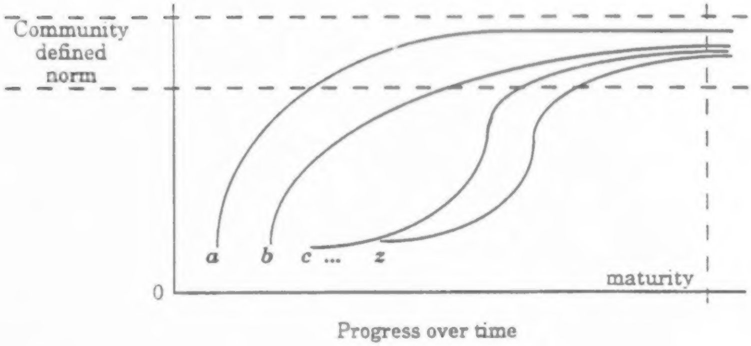
A rather different application of IRT comes from Lynch, Davidson, and Henning (1988). While Kunnan (1990) was interested in variance across items, Lynch, et al., focussed on variance within persons (on a different form of the same UCLA ESLPE examined by Kunnan). Lynch, et al., wanted to determine if variance within persons could also be regarded as unidimensional. It had been determined in several prior studies that variance across items tended to be unidimensional. Both person variance and item variance need to be unidimensional in order for one-parameter Rasch models to be optimally applicable. Like Oltman, Stricker, and Barrows (1990) -- who used a different approach, multidimensional scaling (following Guttman, 1965) -- the evidence obtained by Lynch, Davidson, and Henning (1988) seemed to show that unidimensionality may not be achieved until language learners gain some maturity in the target language. Their conclusion expresses this idea negatively: with reference to violations of unidimensionality, they say that their results seem to support the notion that such violations are more serious at the lower end of the ability continuum (p. 218).

Citing Oltman and Stricker, Lynch, et al. note that the few dimensions detected tend to merge into a larger primary dimension at the upper end of the ability scale (p. 207).

This same observation has been made by Oltman, Stricker, and Barrows (1990) on the basis of a different statistical technique (multidimensional scaling).

Whereas Lynch, et al., studied responses of 678 subjects taking the UCLA ESLPE in the Fall of 1987, Oltman and colleagues studied 53,169 subjects who took the *Test of English as a Foreign Language* in May of 1985. These results give fairly persuasive evidence that whatever factors or dimensions language proficiency may resolve [illegible] probably do vary dynamically over time just as Clifford (1980) and Lowe (1980) predicted they would. In fact, Figure 2 suggests an abstract idea of the sort of thing that appears to be happening with the TOEFL and with the UCLA ESLPE as well. Whereas in the early stages of second language learning, distinct dimensions of listening, writing, and reading ability may be observed (and these may even resolve into further sub-component traits or categories), as learners progress to a more mature, native-like capacity in the target language, it seems that the diverse dimensions (factors, traits, or whatever they may be called) tend to converge to a more unidimensional structure.

Figure 2
Hypothetical Convergence of Arbitrarily Designated Factors or Dimensions Designated a, b, c, ...z (traits, methods, or whatever) of Language Proficiency Viewed Over Time until Maturity is Attained.



A tentative hypothesis may be offered: Perhaps the various dimensions (whether attributed to persons or to items) that are sorted out by language tests (and observed in some detail through multidimensional scaling techniques) tend to converge on some more or less well-determined norm that is defined by the community of users who know and use the target language in question for the sorts of purposes that the language tests inadvertently characterize. There are good theoretical reasons to suppose that some sort of normative convergence must in fact occur in "normal" language acquisition. Whereas learners may vary considerably in the rate and degree of initial success in mastering all of the diverse aspects of a language system, the sounds and meanings of words, the syntax and semantic values of phrases and clauses, not to mention pragmatic applications in experience, must all *tend* toward more or less standardized norms in order for communication to be possible across the diverse members of any given language community. It is precisely in this sense, I believe, that language tests must always to some degree be normative in principle. Criterion-referencing is not ruled out, but it will necessarily be incomplete unless supplemented by norm-referencing (i.e., specifically to the norms of the language community in question). Languages, whatever else they may be, are intrinsically, norms of symbolic behavior. We will return to this idea in section 3 below, but first it may be useful to examine some of the broader research on the measurement of human abilities in order to appreciate better the special role played by language abilities.

(2) Review of Educational Measurement

Modern variants, of the analytic approach typified by the discrete-point foreign language testing of the 1960s can still be found in abundance in the general literature of educational measurement. Kagan (1990) complains about the "atomistic view of effective teaching that emerged from the process-product research of the 1970s" as well as the mistaken notion that a teacher's competency can be defined entirely in terms of a "laundry list of behavioral objectives" (Howey and Zimpher, 1989; Kagan, 1990, p. 419). Of course, a review of the literature shows that the laundry-lists have not been limited to behavioral objectives for teachers but have been extended to every domain of the curriculum and every sort of testing -- including tests aimed at intelligences, achievement, bilingualism, language disorders, etc.

Nowhere is the atomistic, discrete-point approach more apparent than in the literature about how to construct "items." In fact, the analytic, taxonomical philosophy (reflecting little influence as yet from the Chomskyan revolution; e.g., see the numerous references to the taxonomy of Benjamin S. Bloom still prevalent in the literature) continues to hold sway in most educational and psychological testing. For example, Roid and Haladyna (1982) describe "the heart of what is currently known as CR [criterion-referenced] testing" as the notion that "a domain-based interpretation is possible only when a domain or universe of items has been created and the test is based on a sample from this domain" (p. 28). A domain, according to such thinking, is conceived of as a list of potential items from which a sample is drawn in constructing a test. Roid and Haladyna (1982) attribute to Bormuth (1970) the idea that a technology of item writing might "be based on the transformation of sentences into questions" (p. 99). A domain, by this view, is a list of sentences. They acknowledge that the whole idea of sampling from a domain of sentences is susceptible to "serious objections" that arise in connection with "the meaningfulness of definable universes" (p. 34).

There are really two problems here: modern linguistic theory shows that the number of sentences in any given domain of interest for practical purposes is non-finite, and it also shows that any known method of algorithmically generating sentences will produce a great deal of nonsense. Roid and Haladyna (1982), without apparently understanding the linguistic necessities, say "there is a chance for endless mapping sentences, facts, and facet elements, with lack of agreement among developers being a major detriment to progress" (p. 132). The non-finiteness of sentences about any given subject matter renders the idea of a "randomly selected representative sample" uninterpretable, and the abundance of nonsense that would be generated by any known algorithmic procedure makes that approach relatively unappealing. Further, the recommendation (of Bormuth, 1970, cf. Roid and Haladyna, 1982, p. 92) that all possible items in a domain be specified is logically (in principle) unattainable. For these and other reasons, I still believe (cf. Oller, 1979, pp. 32-33) we need to look for an approach to educational and psychological testing that assesses the relative efficiency of a generative system (i.e., the symbolic system itself) rather than attempting to representatively sample from an unattainable listing of an infinitude of demonstrably infinite universes of particular sentences or test items. When the focus is shifted from a list of items (a poor characterization in any case of any non-finite domain of sentences) to the generative basis which underlies the representations that constitute that domain, we have some hope of achieving both reliability and validity. While approaches to educational and psychological measurement have yet to appreciate the purely theoretical implications of the Chomskyan revolution, happily a movement toward more pragmatic, holistic, testing is nonetheless discernible.

Whereas Roid and Haladyna (1982) view individual test items as the "basic building blocks of tests" (p. ix), they implicitly take into account the contrast between (1) discrete-point theory where individual items are matched with some abstract trait and a more pragmatic approach where (2) the tester/teacher thinks in terms of "a theory of the relations between a test and other variables in the real world (a nomological network)" (p. 8). The latter approach would seem to address the fundamental problem of pragmatic mapping (also known as abductive reasoning) to which we return in part 3 below. It is also refreshing to read in Roid and Haladyna (1982) that "testing is viewed as a part of instruction and not a separate operation" (p. 30). In this they follow the lead of people like Eva L. Baker (1980) who argues for a comprehensive "integrating" model of "teaching-learning-assessment" (p. 14) where the various activities are merely viewed from different perspectives, but not as distinct and separate entities apart from the whole context of education. It is the articulation of a theoretical basis for such holistic, nomological, or pragmatic approaches, the author will argue in section 3, that is most needed.

The author agrees with Gardner (1990) who cites Chomsky (1975) in support of the idea that the acquisition of various representational abilities -- though not always the more abstract academic ones that Gardner calls "literacy, numeracy, and critical thinking" -- is natural and normally proceeds without a hitch. "Given environments that are not grossly impoverished, all children will learn how to speak and understand their native languages (and other languages in their surround) with ease and facility; acquire basic understandings of the operation of the physical world (the constancy of matter, the principles of cause and effect); understand key aspects of the social world (the way to convince another individual, the detection of benevolent or malevolent motivation); and use a range of symbolic codes, such as those involved in picturing, gesturing, and making music, in order to express and derive meanings" (pp. 89-90). Following Chomsky, Gardner acknowledges that not only do children normally accomplish such things without special tutelage, but that "adults do not know *how to teach* [his italics] many of the most important forms of knowledge which every normal child acquires" (p. 90.

Gardner in all of his recent writings stresses the partial independence of "intelligences." He says, "While such areas as reading, or studying history, or composing music may well be characterized by stages of competence, the stages found in one domain may have little resemblance to, or correlation with, those regnant in other domains... even in those areas of learning which appear to be universal, all forms of learning do not develop in synchrony. Rather, human beings differ in the manner in which, and the speed with which, they express various mental capacities or 'intelligences' " (pp. 90-91). He points out that learners often exhibit what may be called "U-shaped" growth or learning curves. They seem to acquire a concept but fail to generalize it appropriately to new contexts or over-generalize it to contexts where it does not work. He argues that what is missing in such cases is what he calls "connecting tissue" that would relate abstract symbolic representations to the world of experience more articulately and more completely. In my terms, what is missing is the sort of pragmatic mapping that all genuine learning requires. Too much discrete-point, surface oriented materials passes for curriculum and yet does not achieve much effect. Students remain without the pragmatic linkages to their experience that would make sense of such materials.

Gardner (1990) says that "so long as testing is geared exclusively to 'school knowledge' " -- i.e., the surface-oriented, discrete-point, unintegrated variety -- the "credentials provided by the school may bear little relevance to the demands made by the outside community" (p. 93). To remedy the situation, he is concentrating his efforts on developing "new forms of assessment which are sensitive to particular intelligences and which can document the kinds of learning that take place 'in context' in which students carry out projects of some scope" (p. 104; also see Gardner, 1989; and Gardner and Hatch, 1989). He says that "finding the topic or skill with which one feels 'connected' is the single most important educational event in a student's life" (p. 104; also Gardner and Walters, 1986a).

In coming to his eventual list of seven basic intelligences, Gardner and colleagues examined several sources in the literature: (1) normals (2) pathological and special populations including such cases as autism, savantism, and learning disabilities. Gardner and Hatch (1989) claim that it is possible to escape the biased confines of "linguistic and logical skills" by developing what they call "intelligence fair measures" that "seek to respect the different modes of thinking and performance that distinguish each intelligence. Although spatial problems can be approached to some degree through linguistic media (like verbal directions or word problems), intelligence-fair methods place a premium on the abilities to perceive and manipulate visual-spatial information in a direct manner. For example, the spatial intelligence of children can be assessed through a mechanical activity in which they are asked to take apart and reassemble a meat grinder....Although linguistically inclined children may produce a running report about the actions they are taking, little verbal skill is necessary (or helpful) for successful performance on such a task" (p. 6). Here Gardner and colleagues seem unaware of relevant research by A.R. Luria (1959, 1961, 1979; also Luria and Yudovich, 1959). Luria showed that the integration of verbal skills with certain motor tasks was essential to successful performance of those tasks for children at an early stage of development (e.g., being able to push a button consistently when a green light was on but not when a red light was on).

Serendipitously, in keeping with caveats of pragmatic testing, however, Gardner and colleagues (e.g., Gardner and Hatch, 1989) recommend holistic, highly pragmatic assessment procedures: "even at the preschool level, language capacity is not assessed in terms of vocabulary, definitions, or similarities, but rather as manifest in story telling (the novelist) and reporting (the journalist). Instead of attempting to assess spatial skills in isolation, we observe children as they are drawing (the artist) or taking apart and putting together objects (the mechanic)" (p. 6). Their approach they admit "blurs the distinctions between curriculum and assessment" (p. 5) but this surely we must applaud. It falls in line with recommendations coming from a number of quarters these days for blurring not only the lines between teaching and testing but also between the school, home, and community (Simich-Dudgeon, 1987; and Quintero and Huerta-Macías, 1990).

Parent involvement is stressed by Quintero and Huerta-Macias (1990): they say, "the positive impact of parents' involvement in their children's education is well documented (here they cite among others Simich-Dudgeon, 1987 and Wells, 1986)" (p. 307). They point out that "instructional activities must not only be interactive in nature, but also rich in cultural meanings, comparisons, and critical analysis for making classroom and out of classroom connections" (1990, p. 312). Or, as Freire and Macedo (1987) put it, "the command of reading and writing is achieved beginning with words and themes meaningful to the common experience of those becoming literate, and not with words and themes linked only to the experience of the educator" (Quintero and Huerta-Macias, 1990, p. 42). Or, from a different angle, Smith (1989) says, "individuals become literate not from the formal instruction they receive, but from what they read and write about and who they read and write with" (p. 353). Quintero and Huerta-Macias argue for a "whole language approach" (citing among others Bruner, 1984; Goodman, 1986; and Smith, 1984) they define it: "the whole language approach to language learning emphasizes that language be taught naturally as it occurs within any social environment instead of segmenting it into bits and pieces" (1990, p. 307). They recommend an experience-based approach appealing to the rich existing experiences of the family (Auerbach, 1989).

However, it is important to keep in mind, as Miller (1990) stresses that the broader and deeper view of literacy that whole-language approaches advocate also suggests connections that have too long been neglected: "Literacy viewed from the perspective of communication arising from shared activities with meaningful others cannot be separated from the issues of intelligence, learning, and language...literacy becomes entwined with how and what people know -- with intelligence" (p. 2). When this broader view is assumed, we may hope for better results in education. Quintero and Huerta-Macias (1990) conclude: "In sum, because Project FIEL [Family Initiative for English Literacy] stresses language use in meaningful context, the student's needs, wishes, and past experiences naturally become the teaching methodology, and flexibility of the curriculum is a natural result. Program goals are reached by students, parents, and teachers working together through interaction and learning for real-life needs. Finally, the experience of the project indicates that when social context is attended to in a positive way and the dignity of the learner is upheld, learning occurs" (p. 312).

By using context-rich materials and activities that engage children more fully and challenge their "intelligences" more specifically, Gardner and Hatch (1989) report higher motivation and evidence of a greater diversity of abilities. They report on a study in 1988-1989 with 20 preschool children who were tested on "story telling, drawing, singing, music perception, creative movement, social analysis, hypothesis testing, assembly, calculation and counting, and number notational logic" (p. 8). The authors conclude that only the activities requiring "logical-mathematical intelligence" proved significantly correlated with each other (r = .78, p < .01)." Their analysis, however, may be more detailed than the small number of preschool subjects in their study would justify. In a follow-up with first graders, 15 in all, again the conclusions are perhaps too general to be sustained by the small number of observations involved, but some evidence is provided showing that children do differ in expected ways on the different intelligences posited.

Walters and Gardner (1985) say that "each intelligence" (of the seven Gardner had previously identified) "must have an identifiable core operation or set of operations": for example "one core of Linguistic Intelligence is the sensitivity to phonological features" (p. 4). They say,"While it may well be possible for an Intelligence to proceed without an accompanying symbol system, a primary characteristic of human intelligence may well be its gravitation toward such an embodiment" (p. 5). Of course, if we follow C.S. Pierce, we must suppose that a sign system of some sort is prerequisite to any intelligence whatever. Here is where some additional theoretical development, I believe, is needed.

Another trend in the general educational-psychology literature that corresponds to a move away from atomistic analytic approaches and toward more holistic pragmatic procedures can be seen in studies of language disorders and learning disabilities. Audet and Hummel (1990), for instance, give an interestingly pragmatic analysis of the discourse of a nine-year-old boy diagnosed as language-learning disabled and behaviorally disordered. In general, they followed the discourse analysis procedures recommended by Damico (1980, 1985a, 1985b, and 1991). Although, Adams and Bishop (1990) and Bishops and Adams (1990) did a less fine-grained analysis (see their comparison of their own with Damico's approach on p. 260), like Damico (1985b) they were also able to show substantial reliability for judgments of pragmatic appropriateness. The shared point in all these cases, however, was to give greater attention to pragmatic aspects of discourse (an approach also advocated by Miller, 1990 and by Prutting and Kirchner, 1987).

(3) *Language Proficiency in Relation to a Theory of Intelligence*

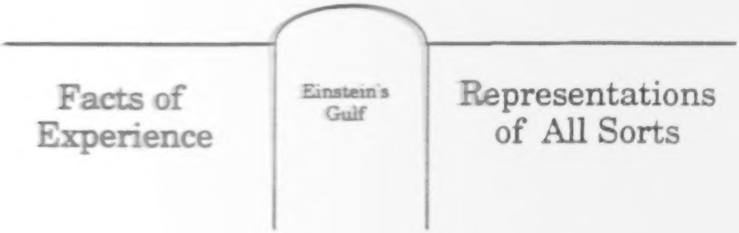
The bulk of the research on intelligence measurement per se is only tangentially relevant to a theory of language proficiency in relation to a comprehensive model of intellect. The IQ measurement research has been limited by its taxonomic character from the beginning and has scarcely begun to consider the full implications of the Chomskyan revolution. The fact is that psychology and psychometrics are yet to feel the force of generative theory. Taxonomic models, e.g., Guilford's "theory of intellect" (1967) and Bloom's taxonomy (1976; also Bloom and Krathwohl, 1977), are not merely out of date, they are either incorrect in fundamental ways, or else, the generative conception of grammar is entirely misguided. At any rate, the taxonomies, when compared against generative theories, cannot compete in scope or power. They are logically too impoverished to even begin to account for the facts of human language ability not to mention other semiotic capacities.

On the other hand, the generative conception of grammar was implicit in much work before the Chomskyan era. Such a conception was apparent in Saussure's advocacy of a general theory of "semiology." Before that, C. S. Pierce [1839-1914], a scientist characterized by Ernest Nagel in 1959 as "the most original, comprehensive, and versatile philosophical mind this country has yet produced," had written the equivalent of 104 volumes of 500 pages each in octavo, focussed primarily on the theory of semiotics. Pierce, more than any other scholar, worked toward a general theory of representations. The essence of Pierce's conception of the relation between language and intellect is suggested by Albert Einstein (1941):

> Everything depends on the degree to which words and word-combinations correspond to the world of impression.
>
> What is it that brings about such an intimate connection between language and thinking? Is there no thinking without the use of language, namely in concepts and concept-combinations for which words need not necessarily come to mind? Has not everyone of us struggled for words although the connection between "things" was already clear?
>
> We might be inclined to attribute to the act of thinking complete independence from language if the individual formed or were able to form his concepts without the verbal guidance of his environment. Yet most likely the mental shape of an individual growing up under such conditions would be very poor. Thus we may conclude that the mental development of the individual and his way of forming concepts depend to a high degree upon language (1941, in Oller 1989, p. 62).

Pierce and Saussure, presumably for similar reasons, agreed in this assessment. Both of them contended that language is the canonical semiotic medium and that by the systematic study of it we should be able to optimize our understanding of representational ("semeiotic," Pierce's term, or "semiological," Saussure's term) processes in general. More recently Noam Chomsky has urged the same program. He wrote in 1972: "One would expect that human language should directly reflect the characteristics of human intellectual capacities" (p. ix).

Figure 3
Pragmatic Mapping of Representations onto the Facts of Experience via Abductive Reasoning



Figures 3-7 elaborate on this central theme. Figure 3 pictures the primary representational problem as outlined in the above remarks by Einstein, and more fully by Pierce in the Nineteenth Century. On the left hand side of the diagram the raw uninterpreted facts of experience are pictured; on the right hand side, representations of them. The question for a theory of intellect is how the connection between the two realms is accomplished. This in a nutshell is the pragmatic mapping problem, or in Pierce's words it is the problem of abductive reasoning. It is construed, in the theory under consideration, to be the primary problem of intelligence.

Einstein described this problem and defined the "gulf" as shown in the following lines:

> ...the concepts which arise in our thought and in our linguistic expressions are all -- when viewed logically -- the free creations of thought which cannot inductively be gained from sense experiences. This is not so easily noticed only because we have the habit of combining certain concepts and conceptual relations (propositions) so definitely with certain sense experiences that we do not become conscious of the gulf -- logically unbridgeable -- which separates the world of sensory experiences from the world of concepts and propositions (1944, in Oller 1989, p. 25).

Readers familiar with Chomsky's work will not fail to see the profound similarity between what Einstein says here and what Chomsky has said many times elsewhere. The idea that true representations are validly connected with whatever they purport to represent, otherwise known as the correspondence theory of truth, is foundational to what Einstein is saying in the immediately preceding quotation. Moreover, it is implicit in many of the remarks of educators concerning the need to relate what is talked about in the classroom to the actual, real-life, real-world experience of students both in and out of the classroom.

Probably the main reason that the Peircean or Einsteinian view of reality has not been more widely accepted by scholars is owing to a peculiar skepticism about our knowledge of the external world that still prevails in much modern thinking and education. MacNamara (1989) shows that modern approaches to human representations often assume an extreme variety of such skepticism. In reviewing a collection of works representing some of the most widely read theoreticians of the present decade (Umberto Eco, Roger Schank, Ray Jackendoff, George Lakoff, and others), MacNamara (1989) complains that "the collection radiates skepticism about the capacity of the mind to know reality" (p. 350). While some of the authors see mental models as mediating between representations and the external world, others see them as being only in contact with themselves. Now it follows that if mental representations have only themselves or other mental representations as their ultimate objects, thinking is quite independent of any external reality, and must be regarded as essentially unrelated to our actions. Common sense and all logic rejects this extreme view. On the contrary, we suppose that people are responsible for their actions in a way that inert objects and unreasoning organisms are not and that the responsibility is based in the linking of representations with corresponding facts that have an independent reality of their own.

When a representation corresponds faithfully to a fact we say that the representation is true of that fact. This is the layman's definition of truth and it does not differ in any essential respect from that of the scientist. However, some skeptics suggest that the very correspondence of a representation with a factual state of affairs is itself a fiction. For instance, Umberto Eco capsulizes this view in his chapter title, "On truth, a fiction" (in Eco, Santambrogio, and Viola, 1988). While C.S. Pierce, whom Eco claims to follow, saw truth as a purely abstract quality of representations (which would give it the same immaterial quality as any fiction -- thus making it fiction*al*), Pierce did not assign any extra degree of reality to material entities so the abstractness of truth would not detract in the least from its reality. On the contrary, while physical things, owing to the laws of thermodynamics come into existence in space and time, grow old, wear out, and are no more, the truth of any representation (e.g., that these words were written by yours truly in Albuquerque, New Mexico, at about 2:25 in the afternoon on August 4, 1991) is an eternal fact. It does not change over time. Therefore, for Pierce, truth was not a fiction, though it has the same abstract *quality* as a fiction. The difference between these views is like that between a libertarian skepticism on the one hand, and a responsible pragmatism (or what Pierce called "pragmaticism" to distinguish his views from those of William James and John Dewey) on the other.

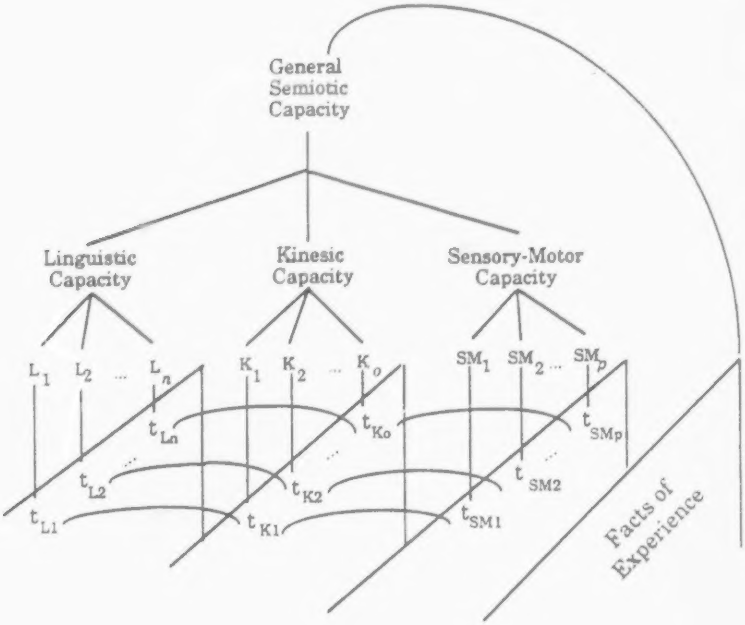
I have mentioned skepticism because it is probably the prevailing view among theoreticians of the twentieth century in spite of the fact that the typical school teacher takes a more realistic approach. For instance, when educators and parents speak of relating classroom activities to the real world, they presuppose that a real world exists and that we have some more or less valid knowledge of it. Therefore, if whole-language, experience-based, socially relevant curricula are actually possible, the extreme variety of skepticism must be wrong.

Figure 4 elaborates on the model by proposing a hierarchy of three distinct kinds of representational capacities: linguistic, kinesic, and sensory-motor. According to Pierce, the language capacity is fully abstract and may be used to represent any imaginable, or even unimaginable idea whatever. We may at least speak of the unimaginably fantastic. The kinesic, gestural, sort of representation is intermediate. It is conventional and arbitrary to some extent, but may also involve iconic (analogical) elements. For instance, a brandished fist suggests more or less iconically the act of punching someone, but it may by convention acquire a rather different meaning -- e.g., it may be a sign of solidarity or brotherhood.

Or consider the fact that Americans and most western Europeans indicate themselves kinesically by pointing roughly at their own sternum (the center of the chest) with the right index finger or thumb of the right hand. Japanese, however, point to themselves by touching or pointing toward their nose with the right index finger, palm turned inward toward the body. Each of these gestures has its conventional aspects as well as its universal basis in the ego-reference point. The latter is not a mere convention since it is physiologically impossible for a perceiver to have any other primary reference point. (Without the notion of one's own self, it would be impossible to credit any other self with existence or to differentiate the self from any other person; see Pierce, in Moore, et al., 1984, pp. 201ff.)

Sensory-motor representations on the other hand are more or less directly, and iconically, related to the facts of experience. Persons skiing down a mountain not only represent the terrain ahead in a continuous flow of images but must also represent at some level body postures and internal commands for motor adjustments in order to control body and skis to accommodate the slope beneath them.

Figure 4
The Semiotic Hierarchy in Terms of Principal Systems of Representation



As Pierce showed with unassailable logic and meticulous phenomenological analysis, sensory-motor representations are analogues, copies, or icons of the facts they represent and, as such they are degenerate. If we look away from an object, its image quickly fades. Details are lost or may be wrongly reconstructed in the mental picture.

Kinesic representations are similar in character, yet may contain an added conventional element. For example, Europeans and westerners in general are apt to point with the index finger to call attention to an object or event. Navajos achieve the same purpose by extending the lower lip. Indexes, a second kind of representational form, are reactionally degenerate. They are not generally so explicit as to rule out the possibility of our noticing the wrong thing which amounts to failing to notice whatever was pointed at. Pierce called this special kind of degeneracy, reactional and distinguished it from the qualitative degeneracy of icons.

Linguistic representations by contrast achieve a higher level of abstraction and a closer approximation to validity. It is true that they must involve icons and indexes to the extent that they are synthetic in character, i.e., to the extent that they inform us about actual experience, but their fundamental character pertains to their abstractness and near independence of anything external to them. While linguistic forms that depend on sensory-motor representations of non-linguistic states of affairs (e.g., factual or fictional contexts), or that appeal to indexical or deictic relations (e.g., pointing or naming or referring) involve the same kinds of degeneracy associated with icons and indexes respectively, the purely semantic values associated with words and propositions are quite impervious to either of those sorts of degeneracy. For instance, our concept of mortality does not deteriorate from one moment to the next in the way that our recollection of a scene does. That is, the semantic value of a word or proposition is not qualitatively degenerate. Nor does our idea of mortality depend on any particular instance of it that might be singled out for attention (e.g., the fact that Socrates died). In fact our abstract concepts (or the abstract meanings of words, propositions, and texts/discourses are not at all reactionally degenerate in the way indexes are. Therefore, Pierce argued, symbols are relatively genuine, i.e., pure and valid by comparison to icons and indexes.

In addition to the fact that linguistic representations are primarily symbolic while gestures have an intrinsic indexical quality in many instances and sensory-motor representations are largely iconic, a few more words need to be said about the three main categories of semiotic systems. Because of their greater abstractness and symbolic character, linguistic representations and their underlying forms embody certain cognitive powers of reasoning that the other two major classes of representations are not capable of achieving. For instance, there is no way that any iconic representation can express adequately the notion that human beings are mortal. Nor is it possible to express that idea strictly speaking in an index or any other sort of mere gesture. An abstract grammatical system capable of expressing a practical infinity of subject-predicate relations, negations, conjunctions of ideas, and the like is required to express fully what is meant by the fact that human beings are mortal or any other similarly complex abstract proposition. However, kinesic and sensory-motor representations also have certain special properties. For instance, an iconic representation, such as a visual representation of a scene, cannot be quite perfectly translated into words. The Chinese aphorism that a picture is worth a thousand words is an understatement. A picture is worth many more than a thousand words. Similarly, gestural systems have unique capabilities. Just as a picture is worth a thousand words, a single look, a facial expression or tone of voice may speak volumes. Affective information, it seems, the emotive side of human experience is far more effectively conveyed in facial expression and tone of voice than it ever could be in words or images alone.

Therefore, each of the three major semiotic systems has its own special capabilities. Still, it must be said that language reigns supreme as commanding the greatest degree of independence from the material world and also, by far, the greatest degree of generality relative to its scope. We cannot visualize, hear, smell, taste, or feel everything we can talk about, nor can we express in paralinguistic mechanisms every idea we can talk about. On the other hand, we can talk about absolutely anything that is conceivable. Anything beyond our capability to represent in some oblique manner in words is simply beyond our conception altogether.

So much for the three general headings under the overall intellectual ability termed "General Semiotic Capacity" in Figure 4. It remains to explain the terms subordinate to each of these. Under "Linguistic Semiotic Capacity," an ability that is believed to be innate and species specific to human beings, come terms that correspond to the grammars of particular language systems, L_1, L_2, through L_n. These systems, to the extent they are not already specified by innate knowledge of universal grammar, must be acquired if they are to be known at all. Each in its turn corresponds then to a class of textual representations in experience, $t_{1.1}$, $t_{1.2}$, through $t_{1.n}$. These terms stand for the texts, for instance, that conform to one's primary language, or second language, and so forth. For monolinguals, there will be no L_2.

The same sort of hierarchical arrangement is hypothesized under the "Kinesic Semiotic Capacity." It too is expected to be largely innate though not entirely species specific to human beings. Again, the universal kinesic capacity dominates (or branches into) a plurality (or at least a potential plurality) of subordinate acquired systems. Each of these subordinate systems dominates a class of texts or representational forms in experience, and these tend to be loosely tied to linguistic texts. For example, English speakers are apt to accompany the statement that a certain person is about "so tall" with a corresponding gesture, palm down, hand extended. A speaker of a different language may use a quite different conventional gesture for the same purpose.

More importantly, research shows that the sequence of gestures is delicately coordinated with the sequence of linguistic forms and meanings. According to research by Condon and Ogston (1971) this is true not only of the speaker but also of the audience to such an extent that their body movements appear to be under the control of one and the same puppeteer.

The case for Sensory-Motor Capacity, if anything, is more dramatic. There is no question that much of our ability to perceive the world and our body as part of it, must be innate (cf. T. G. R. Bower, 1971, 1974; also the Chomsky and Piaget debate in Piatelli-Palmarini, 1980 and comments from the other participants). However, every normal person operates in ordinary experience by so many routines and patterns that it would be impossible to estimate how many distinct sensory-motor systems an ordinary individual possesses. There are sensory-motor programs for almost every imaginable aspect of routine experience, chewing gum, brushing your teeth, grooming in general, dressing, tying your shoes, driving a car, riding a bicycle, playing basketball, going to class, giving a talk, writing a letter, typing one, talking on the phone, etc., and each of these routines is divisible into subroutines of a great variety.

To the extent that such programs can be made explicit as rule-governed systems, they are like grammars of natural languages. They also have their own sensory-motor texts, t_{SM1}, t_{SM2}, and so forth. For instance, our ability to recognize a game of basketball and to distinguish it from a tennis match, or to distinguish either of these from a boxing match, is dependent in part on our knowledge of the corresponding sensory-motor systems. But none of these knowledge systems is the same as an actual game of basketball, or tennis, or a particular boxing match. Yet, the general rule-systems underlying the particular manifest forms (t_{SM}'s in Figure 4) are at least as distinct from each other as are the diverse "textual" manifestations. Sensory-motor texts, in their turn, are also coordinated in ordinary experience in delicately articulate ways with kinesic and linguistic texts.

Because the information processing approach to the development of semiotic systems over time is discussed in Damico and Oller (1991) along with a detailed analysis of some of the empirical evidences in favor of the theory, I will merely summarize those evidences here and will skip over much of the discussion given there (Damico and Oller, 1991) of the theory from an information processing point of view.

Empirical evidence in favor of the theory sketched out includes first, a plausible explanation of our ability to translate information from one semiotic system into another. Each of the universal systems of knowledge (and no claim is made as to the completeness of the ones postulated, only their necessity) though distinct, is related to the others through the domination of the general capacity, and each also subordinates one or more particular systems that are acquired and are to some extent conventional in character. For example, the acquisition of the primary language at once fleshes out the universal aspects of language that are realized in that system and at the same time results in the addition of conventional features that are unique to the primary language. Much the same will be true in the acquisition of the kinesic system that accompanies the first language. Our ability to translate information from one system more or less adequately into another is indicative of the underlying general capacity that connects the different quasi-independent modules or in Gardner's terms "multiple intelligences." We can talk about what we see or describe in words the meaning of a gesture, facial expression, or tone of voice. Or, we can visualize a scene as someone else describes it, imagine a facial expression, tone of voice, or the like based on a linguistic representation. Paraphrase is included as a special case of such translations. We can also paraphrase meanings that have been expressed in a certain surface form by putting them into other surface forms that give more or less the same result. For instance, the statement that "Men are mortal" may be paraphrased by saying that "All humanity must ultimately face death" or that "Mortality is a trait of human beings," etc. Translation across distinct language systems, e.g., "Los hombres son mortales" or "La mortalidad es una de las cualidades de los hombres," or translation into any language or other form that can be imagined, is ample evidence in favor of a general factor of semiotic capacity. Apart from such a general capacity, such translations (even quite imperfect ones, much less fully satisfactory ones) would be inexplicable.

I agree with Roid and Haladyna (1982) as well as Anderson (1972) who recommend the use of paraphrase in the testing of comprehension of prose materials in a school curriculum. Roid and Haladyna (1982) say that "the reason for using paraphrase [in testing] is to ensure that students have truly comprehended the ideas... that they have not just recalled the wording at a surface level" (p. 91). They quote Anderson (1972): "to answer a question based on a paraphrase, a person has to have comprehended the original sentence, since a paraphrase is related to the original sentence with respect to meaning but unrelated with respect to the shape or sound of the words" (p. 92). My point, however, is a little different than theirs as I am stressing the fact that all comprehension of a semiotic sort involves a sort of paraphrasing or translation into a different semiotic medium. This idea comes from Pierce and was viewed by Roman Jakobson (1980) as the special genius of the whole Peircean perspective on semiotics and linguistics. Jakobson commented that "the translation of a sign into another system of signs" as a definition of the process of interpretation was "one of the most felicitous, brilliant ideas which general linguistics and semiotics gained from the American thinker" (p. 35).

Now here is where the theory of Walters and Gardner runs into a difficulty: if there were really independent "intelligences," it should not be possible to translate very well from one to another. They, of course, admit that it is possible to do some such translation and yet at the same time see this as a bit of a "conundrum." They give an example of a non-mathematically inclined child who must master some mathematical principle. They say, after the mathematical approach fails, "the teacher must attempt to find an alternative route to the mathematical context -- a metaphor in another medium. Language is perhaps the most obvious alternative, but spatial modeling and even a bodily-kinesthetic metaphor may prove appropriate in some cases.

In this way, the student is given a *secondary* route to the solution... perhaps through a medium that is relatively strong for that individual" (p. 20). What this potential detour to the difficult mathematical principle shows is that it must be possible to some degree to translate between the different symbolic media. However, they surmise that "there is no *necessary* reason why a problem in one domain *must be translatable* into a metaphorical problem in another domain... as learning becomes more complex, the likelihood of a successful translation diminishes" (p. 20). They assert, "the mathematical principle cannot be translated *entirely* into words (which is a linguistic medium) or spatial models (a spatial medium)" (p. 19). However, no proof of this has been offered, and Peircean theory shows that one of the properties of truly symbolic systems is their relatively perfect intertranslatability. While we cannot translate from an icon to an index, nor vice versa, nor can we always translate from a symbol to either an icon or an index, we can always translate from one symbol to another, and there is no limit to the accuracy of such symbolic translations. Furthermore, all indexes and icons are more or less translatable into symbols, though the reverse is sometimes impossible. How, for instance, would you adequately represent the mortality of human beings by pointing to something in particular? Or what icon would show the full meaning of the symbolic proposition that humans are mortal? On the other hand, a verbal description may suggest an icon just as it may suggest a particular index. In fact, verbal descriptions can literally include icons and indexes within them so as to more or less completely usurp their special representational capacities.

The fact that fairly complex translations are meaningful is demonstrated in the sort of research exemplified by Nolen and Haladyna (1990). They focussed on two types of study strategies that encourage "deep-processing" (their term): *elaboration* (e.g., "figure out how it fits in with what you learned in class") and *monitoring* ("asking yourself questions while you read to make sure you understand") (p. 117). They argue that "if students think the teacher wants them to understand material and relate it to their own lives, as well as to think creatively and independently about it, they will come to value strategies (like monitoring and elaboration) that lead to those goals" (p. 119). Now if translation of the sort that takes place between distinct semiotic media were not fairly good, it is difficult to see how "deep-processing" would relate to all of the diversity of concepts, illustrations, photographs, texts, experiments, etc. that constitute the curricular bases for learning about science. In fact, the whole thesis of experience-based, socially relevant, whole language education, is grounded in the implicit assumption that meaningful connections and translations across distinct semiotic media are not only possible but more normal than the traditional analytic separation of those media into separate and independent categories.

Another evidence of the connectedness of the various disciplines summed up in Gardner's terms "literacy", "numeracy", and "critical thinking" (Gardner, 1990) is seen in a rare longitudinal study by Benbow and Arjmand (1990) involving 1,247 persons initially identified in the seventh or eighth grade as "mathematically precocious". These individuals were observed again after they completed college to identify factors that contribute to high achievement in mathematics and the sciences. In addition to finding that a high SAT score at age 12 was a good predictor of subsequent performance (however, a mediocre or low score did not yield much predictive value), the authors (Benbow and Arjmand) confirmed the observation of Walters and Gardner (1986a) that there was typically some "crystallizing experience" (event or persons) that contributed to the educational development of the high achievers (p. 437). Two observations are suggested here: first, that testers cannot rely on negative evidence as much as positive evidence of abilities, and second, that influence stemming from interpersonal relations (a mentor or encourager) may have a profound influence on mathematical or scientific achievement. Now this last outcome would seem to be excessively unlikely if the separate "intelligences" labelled "interpersonal" and "logical-mathematical" were truly quite independent. They have to be related via some form of intertranslatability.

The semiotic model under consideration (Figures 3-5) also enables us to make certain distinctions that are, it would seem, critical to any theory of intellect that aims for explanatory adequacy (cf. Chomsky, 1965). For instance, we may distinguish innate from acquired knowledge. Innate knowledge is that which is present before any experience occurs, or which is triggered by experience and matures more or less automatically and somewhat independently of experience. Even sensory-motor systems have their noteworthy conventional aspects. For instance, to take a trivial but suitable case for the sake of illustration, in one culture it is customary for automobiles to drive on the right hand side of a roadway while in another motorists stay to the left. If it is hypothesized that conventional aspects of the various semiotic systems in question must be acquired, this sort of acquired knowledge will be distinguished from innate knowledge to the extent that the former is a product of experience involving the senses. It is suggested that information from the sensory-motor system passes to consciousness where the sensory-motor texts (i.e., sequences of sensory-motor images) are interpreted. As they are understood, and just to that extent, they are passed through various stages of memory more or less distant from consciousness. The depth of the comprehension in question will determine the degree of impact on semiotic systems. It is hypothesized that the acquisition of grammar is a process of comprehending a particular kind of texts so as to develop the sort of intuitive feel which constitutes knowledge of a language. By this reckoning, the acquisition of a particular grammar is a process of comprehending texts in that language at a sufficient depth so as to acquire the conventional aspects of the grammatical system.

Contrary to a lot of recent speculation about non-primary language acquisition (e.g., Gregg, 1988), the theory under consideration hypothesizes that non-primary language acquisition will proceed in a manner much like primary language acquisition except for the fact that acquisition of a second language will benefit greatly (and suffer minor interferences from) the prior acquisition of the first language (Asher, 1969; Asher and Price, 1967; Asher and Garcia, 1969). Similarly, the acquisition of a third language will benefit (mainly, and suffer but little) from the first and second, and so on. The fact that non-primary language acquisition usually falls short of the mark achieved in primary language acquisition (Gregg, 1988), it is supposed, should be explained not by positing a radical difference in the physiology (Scovel, 1988) or even the internal strategies of the person involved in one or the other task (Selinker, 1972), but by noting the radical differences across the two cases in access to target language texts and the relative motivations to comprehend and produce them (Brown, 1973; Schumann, 1975; Vigil and Oller, 1977).

In the primary language situation, the person doing the acquisition is under incredible community pressure to conform to the norms of the primary-language. A child who persists in non-conformities will be ostracized or punished in ways that border on cruelty while the one who succeeds in overcoming them will be rewarded by all the privileges of membership in a community. For any one other than a child acquiring a non-primary language, no similar pressures or rewards are likely to be experienced (cf. Brown, 1973; Schumann, 1975; Vigil and Oller, 1976; etc.). Exceptional cases, where non-primary language acquisition succeeds in fairly dramatic ways are precisely those cases where access to target language texts and susceptibility to pressures and rewards are both provided for. For instance, the person who marries across language boundaries and then moves to the country where the non-primary language predominates is far more apt to achieve native-like ability in the non-primary language than someone who merely takes a college course in that language. In fact, we are inclined to suppose, along the lines of Vigil and Oller (1976) that continuing progress toward native competence in any language is much more a function of internally defined motives and sensitivities than it is a function of methods of teaching or modes of exposure. Clearly access to pragmatically rich and meaningful texts in the target language is requisite, but insufficient by itself. Motivation to conform to the communal conventions of the target language system is also required.

The hierarchical model under consideration not only supports the kinds of theoretical distinctions that are required in practice, e.g., the distinction between innate and acquired knowledge, consciousness and memory, memory and grammatical knowledge, grammar and text, text and comprehension, comprehension and production, primary and non-primary language acquisition, etc., but it also suggests some fairly explicit hypotheses about relationships within the proposed hierarchy that are immanently susceptible to empirical testing.

Since linguistic representations are the most abstract ones considered in the model, it follows that the primary language is the most likely basis for the development of general semiotic capacity. Here I differ some with Walters and Gardner (1985, 1986a, 1986b). They seem to view "logical-mathematical intelligence" as distinct from "linguistic intelligence." But, it has often been observed that logic and mathematics involve kinds of reasoning that are parasitic and derivative being entirely dependent upon language (Pierce, in Hartshorne and Weiss, 1931-1935; Lotz, 1951; Church, 1951; Russell, 1919). Einstein alluded to the closeness of the relationship between language development and cognitive growth in general in the remarks quoted above. It was a point developed further by Vygotsky (1934, 1978), Piaget (1947), Luria and Yudovich (1959) and Luria (1961).

Further evidence may be seen in the remarkable accomplishments of deaf children with hearing parents. In cases where the children, for whatever reasons, are deprived of access to visual sign language they face a language acquisition problem far more difficult than that of the hearing child. Such children, it seems, face special cognitive difficulties that only the acquisition of a fully developed language system will enable them to overcome. Typically this is accomplished through a natural visual-manual sign system such as American Sign Language (cf. Lane, 1984; Wilcox, 1988). (An interesting aside concerning such signed systems is that the primary role of language is assumed by gestures of the hands and body while the paralinguistic role of kinesics is taken over by speech and voice mechanisms.) Deaf children deprived of manual/visual sign system and forced to acquire speech directly are placed at a serious disadvantage (Lane, 1988). The difficulties they face in cognitive development across the board are predicted by the hierarchical model under consideration. It follows that if children are deprived of full and rich primary language system that is accessible to them in terms of their sensory-motor system, they will suffer consequences of this lack throughout the cognitive hierarchy and especially in areas that depend on communication, e.g., social development.

Moreover, children who acquire some ASL and are then taught Signed English (SE), an artificial system invented by hearing persons to correspond to English lexicon, syntax, and so forth, are apparently in the position of persons trying to acquire a second language system. In this instance, however, the system is artificial in a variety of ways. For instance, in theory SE gives equal emphasis to stressed and unstressed morphological and lexical elements. In this respect, and others, it is somewhat like Morse Code or even Pig-Latin. Unlike ASL, SE is a largely dependent system. Therefore, when deaf children de-emphasize or omit redundancies of English structure, e.g., the "-ing" of present progressives and the like, they are making natural modifications in surface forms of signed texts that would conform to more normal expectations about universal grammar.

Another hypothesis that is suggested by the theory under consideration is that neighboring elements of the hierarchy are more apt to influence each other than distant ones. For example, the primary language would have greater impact on second language acquisition than on third. The second similarly would be expected to influence the third, even more than the first language would, and so on. Again, experience of polyglots bears this out. Typically, "padding" (a term from Newmark, 1966, i.e., the use of known language forms in place of target language forms) is usually from the most recently acquired language rather than from any other.

Following out the same idea, transfer in general would be expected to occur from the more developed systems to less developed ones. For example, the primary language would be expected to influence a non-primary language rather than the reverse. The situation would be altered in favor of the non-primary language at just the point where the person in question achieved greater proficiency in the non-primary system. However, at just that point, the non-primary system would be promoted to the status of the primary system and the former primary system would presumably be demoted to a secondary status.

Another consequence of the postulated hierarchy is that distinct representational systems provide the means in some cases for comprehending what would otherwise be incomprehensible. For instance, a discourse in a target language that might be entirely incomprehensible if one had to rely on knowledge of that particular language alone can be made comprehensible if one has access to a translation provided in some other semiotic system. In normal language acquisition, e.g., primary language acquisition, as has often been pointed out (Macnamara, 1973, 1982) meanings of surface forms are often contextually obvious when those forms are being acquired (Krashen, 1985). The child first understands the context, e.g., by representing it in a comprehensible sensory-motor form, and subsequently becomes able to understand the utterances associated with the context. In non-primary language acquisition, wherever it succeeds, a similar scaffolding is often provided. It may be presented in some dramatization, in a film, or it may be presented through a translation, literally, into a language that the subject already knows.

By this line of reasoning, Krashen's input hypothesis (Krashen, 1985) is vindicated (Oller, 1988). The input hypothesis in its most basic form says simply that language acquisition progresses as the acquirer comprehends texts that are a little beyond his or her current level of development in the target language. Spolsky (1985) and Gregg (1988) have contended that the input hypothesis is either false or trivially true. If it means we must understand what is beyond our understanding, it is false. If it means merely that we must comprehend in order to learn, it is trivially true. However, the theory we are advocating here disposes of both of these interpretations. We do indeed understand representations (target language texts) beyond our reach in one system (namely the target language) by appealing to representations in another semiotic system. The one provides an interpretation of the other. Therefore, because of the intertranslatability of semiotic representations, the input hypothesis remains viable.

Cummins (1976) proposed the threshold hypothesis, an idea that relates to the impact of bilingualism, or more specifically adding a second language, on cognitive development. Subsequently (see Cummins, 1984, pp. 107-108) he modified his hypothesis and extended it. The threshold hypothesis suggests that the child's starting level of proficiency in one or both languages may be an important mediating variable in avoiding a burden in becoming bilingual or in benefitting from bilingualism once achieved. There are actually two thresholds being proposed.

On the low end, it is claimed that a child may have to achieve a certain minimal level of proficiency in one or both languages in order to avoid deficits. In other words, if the child falls below threshold in both languages, presumably it will be difficult or even impossible for that child to benefit from instruction in either language. Further, it follows that a child who has not acquired threshold level in the primary language will only receive an unnecessary additional burden by being instructed in a second language. Therefore, the lower threshold is presumably important in the determination of when instruction might be beneficially introduced in a non-primary language.

At the other end of the scale, a high threshold is also posited. In order for a bilingual child to experience the expected benefits of bilingualism, e.g., greater ability to appreciate and utilize symbols and greater "metalinguistic awareness," i.e., ability to appreciate the arbitrariness and conventionality of linguistic symbols, the child must have surpassed the high threshold presumably in one or both languages.

Admittedly, the idea of one or more thresholds is loosely stated, but the research seems to support it (Cummins and Mulcahy, 1978;

Duncan and DeAvila, 1979; Hakuta and Diaz, 1984; Kessler and Quinn, 1980). In fact, as Hakuta (1986; also see Lambert, 1975) has shown, there is a long history of debate concerning the deleterious versus beneficial effects of bilingualism. Formerly, especially in the U. S. there was a widespread prejudice against "bilingualism" based on research showing that minority language children got low scores on IQ tests. It scarcely occurred to the persons interpreting the research that the IQ tests were mainly measures of English language proficiency -- something that the minorities in question had not yet had the opportunity to acquire.

The main point here, however, is that the hierarchical model under consideration explains the available evidence concerning the threshold hypothesis and provides a convenient framework within which to understand the interrelationships of semiotic systems in general. Within a hierarchical model, the threshold hypothesis can be incorporated and elaborated in terms of transfer and interference and in terms of a more explicit theory of the role of language proficiency in relation to cognition in general. Bilingualism and indeed multilingualism deserve special consideration since they are bound to play a central role in the education of minorities. Moreover, the elaboration suggested by the theory under consideration is compatible, it seems, with the course that Cummins (1979, 1983a, 1983b) has begun to develop in terms of the CALP/BICS distinction.

In response to consideration of the possibility of a general language proficiency factor, Cummins (1979) hypothesized a distinction between what he called cognitive academic language proficiency (CALP) and basic interpersonal communicative skills (BICS). This idea was appealing inasmuch as most any educator who has dealt with bilingual or multilingual contexts has observed ample evidence in its favor. A child that gets along satisfactorily on the playground, where cognitive demands are presumably lessened by the immediacy of physical and social context, may encounter difficulty in the classroom when it comes to reading, writing, solving word and math problems, and in general interacting on a more abstract level. The child may have adequate BICS without sufficient CALP. This distinction is reminiscent of the sort of thing Gardner (1990) says in reference to representational systems that seem to be naturally acquired versus ones that need special "tutelage" -- especially, "literacy, numeracy, and critical thinking" -- the sorts of things that Cummins would group under CALP. Cummins (1983c), however, unlike Gardner and colleagues, clarified that he did not intend to argue that the two kinds of ability were unrelated, but rather that they were apt to appear as such at the surface. To illustrate he adapted an "iceberg" model (from Shuy, 1978, 1981) where the two visible points, CALP and BICS, were clearly distinct, but were joined below the surface in what he called "common underlying proficiency" (cf. Cummins, 1984, p. 143).

There was a further implication that the two kinds of ability might be developed in somewhat different contexts and perhaps using distinct strategies. Cummins (1983c) quoted David Olson (1977) who said:

> ...language development is not simply a matter of progressively elaborating the oral mother tongue as a means of sharing intentions. The developmental hypothesis offered here is that the ability to assign meaning to the sentence per se [as in a written text], independent of its non-linguistic context, is achieved only well into the school years (p. 275, cited by Cummins 1983c, p. 116, our interpolation).

What Cummins and Olson apparently intend to emphasize is the greater degree of inference required to link up a written text with its author's intended meanings than is required in the case of an interactive discourse in the here and now. The latter, presumably the typical context of the exercise of BICS, is less cognitively demanding, *ceteris paribus*, than the former, a typical context for the use of CALP.

Within the more elaborate Peircean perspective proposed here, Olson's phrase "independent of its nonlinguistic context" might be reformulated as "without firsthand access to its nonlinguistic context." This seems to do no violence to Olson's intention, nor Cummins application of the idea in reference to CALP. However, it is a necessary modification if Pierce's foundational claim that all interpretation is translation from one form of semiotic representation to another. This sort of translation is not viciously circular only because sensory-motor representations enable the investment of all other sorts of representation with material (non-empty) content.

However, strictly speaking, there is no such thing as a meaningful "sentence" without a "nonlinguistic" context. With that in mind, we assume that Olson and Cummins might accept as a friendly amendment to their ideas the interpretation that CALP (or in Olson's case, literacy) requires a larger inferential leap from the perceptible form of a representation (a written text in the case under consideration) and an appropriate interpretation that associates it with experiential context. Failing this, it would have to be argued that a representation which has no inferential relation to any experiential context whatever is necessarily meaningless. It is entirely uninterpretable (cf. Einstein, 1944, in Oller, 1989, p. 25, paragraph 3.13; and Pierce, pp. 99-105 in Oller, 1989).

How then can the CALP/BICS dichotomy be understood within the proposed hierarchical model? The overlapping part of the iceberg beneath the surface would be explained in part as the general factor of language proficiency which incorporates whatever aspects of general intelligence are necessary to that proficiency. For BICS, also, it is clear that the utilization of both sensory-motor information and linguistically coded representations simultaneously would require a pragmatic linking that could only be accomplished by access to general semiotic ability. However, with BICS, sensory-motor information is immediately accessible to aid the pragmatic linkage.

In the exercise of CALP, on the other hand, say in reading an unillustrated text, e.g., that which appears on this page, any necessary supplementary sensory-motor representations would have to be supplied by the reader. This is a more difficult semiotic task. It requires a higher degree of inference based on a more abstract semiotic system, namely a linguistic one, from which the sensory-motor type images must be inferred where they are needed. The move from graphological representations to a more abstract linguistic form is already a difficult inferential process (reading), and the absence of sensory-motor images that might give some clue concerning reference, deixis, and the whole pragmatic mapping process involves another complex of inferences.

Thus, CALP, with its special emphasis on literacy and abstract reasoning would presumably require the development of reading and writing skills in the primary or some non-primary language. Whereas BICS might benefit indirectly from such a development, literacy and specialized abstract reasoning skills, e.g., ability to do arithmetic leading on to higher mathematical skills, would not be necessary to BICS. To this extent, BICS and CALP are usefully distinguishable which suggests an important amplification of Cummins's threshold hypothesis -- one that he has commented on (Cummins, 1984, p. 117).

The initial distinction between "surface fluency" and "conceptual-linguistic knowledge" Cummins attributes to Skutnabb-Kangas and Toukomaa (1976). They, no doubt, were influenced by the distinction between "surface" structure and "deep" structure from Chomskyan linguistics. The idea was that a child might develop quite a lot of routine facility with greetings, leave-takings, playground games, and the like, and still fall short of the level of language proficiency and concept development necessary to reading, writing, and doing arithmetic (or as Gardner, 1990, terms them "literacy", "critical thinking", and "numeracy"). Therefore, a child might appear to do well at conversation but fail at school (Olson, 1977).

The low threshold for language skill, then, might be construed as a completely general requirement applying as much to monolinguals as to multilinguals. Presumably this same notion was what another generation of specialists in another paradigm meant by "readiness". The higher threshold too would have a more general interpretation in this context. Presumably "metalinguistic awareness" is merely another way of referring to what another generation of psychologists and educators called "learning to learn" or "talking about talk," etc.

Finally, there is also a parallel with the traditional distinction between "language disorders" and "learning disabilities" where the former have been defined more in terms of surface language problems (sometimes even speech difficulties per se) and the latter in terms of deeper conceptual difficulties -- "neurological" deficits (see Coles, 1978; Cummins, 1986) or, more recently, "inefficiencies" (Swanson, 1988). Damico (1985b) has argued that traditional tests of language disorders have tended to focus on surface forms of language while definitions of learning disabilities have been defined, to the extent they have been defined at all, in terms of deeper conceptual problems. Again, something like the BICS/CALP distinction appears. It is a virtue of the proposed model under consideration to be able to incorporate such distinctions and to elaborate upon them in intuitively appealing ways.

Table 1
The Seven Intelligences

Intelligence	End-States	Core Components
Logical-mathematical	Scientist Mathematician	Sensitivity to, and capacity to discern, logical or numerical patterns; ability to handle long chains of reasoning.
Linguistic	Poet Journalist	Sensitivity to the sounds, rhythms, and meanings of words; sensitivity to the different functions of language.
Musical	Composer Violinist	Abilities to produce and appreciate rhythm, pitch, and timbre; appreciation of the forms of musical expressiveness.
Spatial	Navigator Sculptor	Capacities to perceive the visual-spatial world accurately and to perform transformations on one's initial perceptions.
Bodily-kinesthetic	Dancer Athlete	Abilities to control one's body movements and to handle objects skillfully.
Interpersonal	Therapist Salesman	Capacities to discern and respond appropriately to the moods, temperaments, motivations, and desires of other people.
Intrapersonal	Person with detailed, accurate self-knowledge	Access to one's own feelings and the ability to discriminate among them and draw upon them to guide behavior; knowledge of one's own strengths, weaknesses, desires, and intelligences.

BEST COPY AVAILABLE

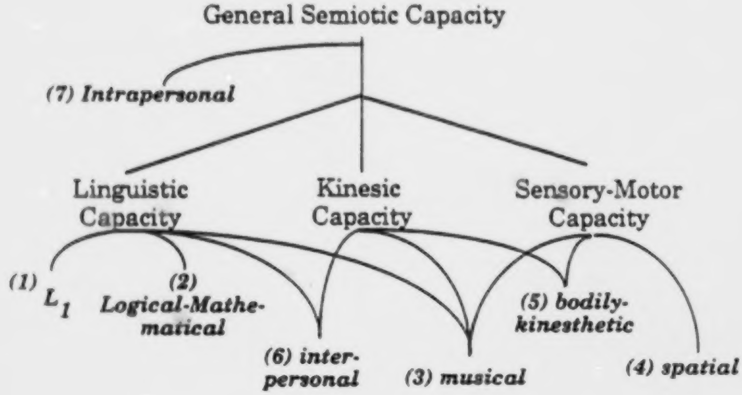
To see better how the proposed hierarchy works in practice, and also to show how it can be used in the evaluation of other theories of intelligence, it may be useful to pause to examine more closely the model proposed by Gardner (1983, 1989, 1990) and colleagues (especially, Gardner and Hatch, 1989; Walters and Gardner, 1985, 1986a, 1986b). Table 1 gives a list of the seven "intelligences" that Gardner sees as somewhat independent of each other and yet as capable of characterizing of the sorts of individual configurations of abilities that he believes necessary to a more adequate conception of intelligence. While Gardner and colleagues speak as if their categories of "multiple intelligences" were thoroughly independent, they are upon examination hardly self-contained, independent modules, but rather complex composites of semiotic capacities in each case. Perhaps they are quasi-modular in character, but it is difficult to see them even in that way. Nevertheless, for the sake of demonstrating the intrinsic compatibility of the quasi-modular semiotic hierarchy I have been discussing here (Figure 4 above especially), I will fit Gardner's categories in as shown in Figure 5 and will discuss them one-by-one in terms of the analysis given by Gardner and Hatch (1989) as well as my own semiotic characterization of their categories.

The first category is what they call "logical-mathematical intelligence" which they describe (see Table 1 above) as pertaining to a "scientist" or "mathematician". It is generally agreed by professional logicians and mathematicians (who have gained some awareness of linguistics) that logic and mathematics are both parasitic and derivative fields of study entirely dependent on human language abilities at a deep level. Therefore, I have placed Gardner's first "intelligence" as a node subordinate to the universal deep language system that is postulated to underlie all abstract symbolic systems as well as natural languages.

Gardner's second category, "linguistic intelligence" characterized in the special proclivities of a "poet" or "journalist" I have associated with primary language ability in the semiotic hierarchy. Gardner and Hatch (1989) give no indication that they have in mind any sort of polyglot, so I do not relate their category directly to the deeper level of universal language ability. That deeper level, I suppose, must undergird all abstract symbol systems such as mathematics, logic, and musical notation, as well as the abstract symbolic aspects of map making, diagramming, illustrating, and in general all forms of what Pierce called "abductive reasoning" (or what I term "pragmatic mapping"; as diagrammed in Figure 3 above).

Gardner's third category, "musical intelligence," as shown in the special abilities of a "violinist" or a "composer," I would place under the sensory-motor class of representations but with special connections to deep language abilities and to kinesic abilities. While a violinist might not be a reader of musical notation, this is unlikely, and a composer certainly would be a reader of music -- hence the connection with the abstract deep language node. In addition, a composer or a violinist would also be apt to understand the sorts of special gestural systems used by conductors (though neither of them might be conductors, a composer would be likely to have the capacity to conduct one or more musicians in performing his or her music) -- hence, the connection with the kinesic (significant gestural) node.

Figure 5
The Semiotic Hierarchy with
Gardner's Seven Categories ("Multiple Intelligences")
Added to the Picture



The fourth kind of intelligence, "spatial," as represented in the special skills of a "navigator" or "sculptor" seems remarkably broad. Surely it covers a multitude of abilities. Among them would have to be found the sensory-motor elements pertaining to perspective and movement in time and space as well as a keen sense of proportion bordering on the mathematical. For the navigator, mathematical skills would surely come into play. For this reason, the "spatial intelligence" is connected both to the sensory-motor node and to the deep language node.

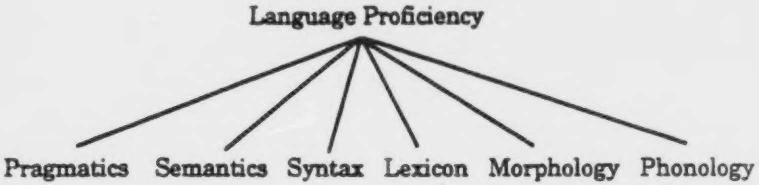
"Bodily-kinesthetic intelligence," Gardner's fifth kind of intelligence, as seen in a "dancer" or "athlete" suggests a multitude of connections as well. If the dancer is a person who understands choreography or if the athlete understands demonstrations of various performances (e.g., how to serve a ball in tennis or how to do a single-leg sweep in wrestling), an implicit comprehension of diagrammatic illustrations would probably come into play. Therefore, I have shown connections to the kinesic node as well as the sensory-motor, but no doubt if coaching comes into the picture, the language node should be connected as well.

The sixth category, "interpersonal intelligence" as seen in a "therapist" or "salesman" suggests again an interesting composite of abilities. Since "moods, temperaments," etc. (as suggested in the descriptor of the category) are discerned largely through kinesic and paralinguistic systems such as gesture, tone of voice, facial expression, and the like, the primary connection would be with the kinesic node. However, to the extent that all sales' pitches tend to rely on linguistic as well as other representations, at least the primary language system would come into play. Since Gardner and Hatch give no indication that the salesperson or therapist they have in mind is a multilingual, connections to languages other than the primary one are not shown, but a polyglot would no doubt have them. Therefore, it is clear that this module of "intelligence" would probably be heavily contaminated by one or more verbal components.

The seventh category is the most problematic of all. Gardner calls it "intrapersonal intelligence" and suggests that it is the ability to understand one's own abilities. The sort of person having this particular constellation of gifts is not only, we may suppose, a rare bird, but one who knows even more about him or herself than the people who are looking for him or her. That is to say, a person who understands his or her own abilities in the way described knows a good deal more than the measurement specialists do. This category, however, I suppose would have to be linked directly to the deepest level of the semiotic hierarchy since it implies knowledge of all the nodes beneath it and of their interconnections. This final observation concerning Gardner's system also sums up my basic objection to it: the interconnections that must be posited if we are to understand how the various modules relate are missing. The sort of semiotic hierarchy that I am proposing here, however, would supply at least some plausible alternatives for such connections.

One of the most difficult things to see about language proficiency is that it may (perhaps must or at least ought to) be conceptualized in a considerable variety of different but mutually compatible ways. Walters and Gardner (1985) assert that "a particularly high level of ability in one Intelligence, say mathematics, does not require a particularly high level of ability in another Intelligence, like language or music. This independence of Intelligences contrasts sharply with traditional measures of IQ that find high correlations among test scores" (p. 13). I agree in large measure with what they are saying provided we modify the word "independence" to "quasi-independence" or something of the sort.

Figure 6
Language Proficiency Viewed as a Composite of Domains of Grammar



With respect to language proficiency per se, it is possible to think in terms of the various components of grammar (Figure 6) that constitute it in theory, or we may think of language proficiency in terms of the traditional skills (Figure 7). Or, we may choose any number of other angles or combinations of them. What is difficult to see is that these are not incompatible ways of viewing the phenomena of interest -- merely different ways. If we focus on primary language ability as represented in Figure 4 above, that portion of the diagram might be amplified as shown in Figures 6 or 7. In Figure 6, language proficiency is seen as divisible, more or less, into domains of grammar. Pragmatics may be defined as pertaining to those aspects of meaning that have to do with actual, particular, concrete contexts of experience. Semantics embraces those aspects of meaning that are virtual, universal, or abstract. Syntax is concerned with the sequential or simultaneous arrangement of categories of grammar into texts. Lexicon comprises those inventories of elements that are acquired as whole units, e.g., words, idioms, pat phrases, verbal routines, and the like. Morphology in English is a question of inflections, e.g., pluralization, tense and number marking on verbs, etc., and derivations, e.g., adding a morpheme to make a verb of an adjective, e.g., "real" plus "-ize" to get "realize," and so forth. Phonology is a matter of determining the surface forms of phonemes, syllables, lexical items, and larger units of structure.

Figure 7
Language Proficiency Viewed as a Composite of Quasi-Independent Skills

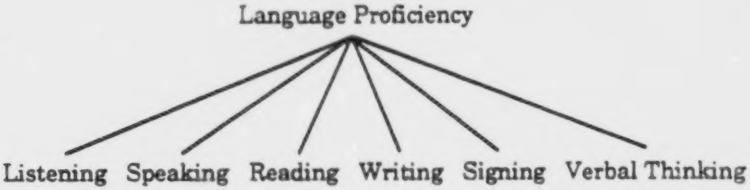


Figure 7 shows a similar breakdown with reference to skills such as listening, speaking, reading, writing, and verbal thinking. It may be argued without risk of contradiction that such hypothetical domains of structure, or distinct skills, are as valid as the theories upon which they are based. However, such divisions can never be finally determined anymore than Immanuel Kant could determine once for all the ultimate categories of reason. As Pierce, Einstein, and others have shown, such categories are intrinsically arbitrary and cannot be finally fixed or completely determined by any amount of empirical research (see especially Einstein, 1941, 1944, and Pierce, 1878, 1906). While it may be possible to fix upper and lower limits within which the simplicity/complexity of the model must fall, its specifics will apparently always retain a substantial arbitrariness nonetheless.

For instance, there is no conceivable argument that would prove either of the componential breakdowns of Figure 6 or 7 to be intrinsically superior to the other. For one purpose one model might be preferred, for some other purpose, another. What is more, many other componential models may be conceived. For example, modes of processing (productive versus receptive) may be distinguished, modalities of processing (articulatory/auditory versus visual/manual), stages of processing (consciousness, short-term, long-term memory), etc. In principle, there are an infinite variety of possible componential models. The answer, therefore, to the advocates of multiple intelligences (e.g., Gardner, Walters, and other collaborators) is that there is no single arrangement that will be completely satisfactory. Within the proposed hierarchy, this fact can be construed as a natural outcome of different ways of combining and/or parsing up various of the proposed elements.

While it was long maintained that cognitive development may be hindered by becoming bilingual, the evidence clearly points in the other direction (cf. Hakuta and Diaz, 1984; Cummins, 1984, 1986; Hakuta, 1986). Dabbling in non-primary language acquisition may have little or no impact on intellect, but the acquisition of a second or third or fourth language to a substantial degree of proficiency is apt to result in significant, though modest, cognitive gains. In particular, the evidence seems to suggest that bilinguals achieve some kinds of flexibility in reasoning and a capacity to appreciate certain kinds of abstract relations that might remain outside the reach of some monolinguals. This result (see the research cited above with reference to the "threshold" hypothesis), is predicted on the basis of the hierarchy under consideration.

Moreover, as in the case of the threshold hypothesis, a more general hypothesis is suggested. If bilingualism contributes to mental growth only after some threshold is passed, it follows that simply attaining proficiency in one's primary or native language must be important to normal mental maturation. Further, if language is a window through which researchers may get a fairly clear look at the mind, a thesis Chomsky has been pushing lately, it follows that the development of language proficiency must be linked to normal cognitive development. Putting this hypothesis in its most general form (Oller, 1991) following Pierce, Einstein, and others, it is possible to predict that the normal development of deep semiotic abilities must depend in subtle ways on the development of the primary language. This has been demonstrated above in part by the differentiation of iconic, indexical, and symbolic representations. Because of its greater abstractness (i.e., symbolic character), language has certain capabilities that the other representational systems lack. Among them is the potential for deep level semantic representations that are quite abstract (i.e., relatively uncontaminated by the two kinds of degeneracy associated with icons and indexes). As a result, only deep language ability is logically a medium that might serve for the development of the most general sort of intelligence. For an elaboration of this idea and a content analysis of so-called "non-verbal" IQ tests showing that they require such deep propositional or semantic reasoning, see Oller (1991).

While it may be possible for deep semiotic abilities to be developed to a high degree with reference to some other manifest form, say, sensory-motor representations, since linguistic representations achieve a more complete level of logical abstractness and conventional arbitrariness, it seems likely that in normal human beings language development in all of its diversity is the fulcrum on which intellect attains its greatest leverage. It also follows that language abilities will tend toward the center of any definition of human exceptionalities ranging from giftedness in all its varieties to disabilities of all types.

(4) Recommendations for Testing (and Teaching) LEP Students

Cummins (1986) writes, "Historically, assessment has played the role of legitimizing the disabling of minority students. In some cases assessment itself may play the primary role, but more often it has been used to locate the `problem' within the minority student..." (p. 29). This process may not have been intentional, but the effect has been summed up by Chase (1977) in a single phrase. He called it "the biologizing of social problems" (cf. Coles, 1978, for concurrence).

Not to deny the fact that some children may indeed have genuine "neurological" or other "deficits" or even "abnormalities," Cummins still contends that the medical "diagnosis/prescription" paradigm has seduced a whole generation of educators and clinicians, and that in many cases children from minority language backgrounds have been ludicrously over-represented in deficit categories (e.g., see Ortiz and Yates, 1983). It is the purpose of this section to discuss these facts in light of the proposed model of semiotic abilities and to show some of the ways that the whole process of assessment might be upgraded and set on a path of self-correcting research and progressively greater adequacy.

It is difficult to over-estimate the pervasive influence of analytic, discrete-point thinking in the study of exceptionalities. Its main manifestation is the search for specific, particular, unique sources of difficulty in individual cases. Swanson (1988), for instance, stresses the aim of the learning disabilities paradigm to achieve "specificity" (p. 197) -- a concept that is elaborated throughout his informative article. This means focussing on "specific mental processes" in instructional remediation and determining unambiguously that "the process under investigation is responsible for performance" (p. 200). The idea of a "generalized deficit," he says, "undermines an important tenet of the field" (p. 197). He complains that "there is a lack of theoretical integration in the choice of measures in subtyping studies, and non-operational definitions of LD exist (Shepard and Smith, 1983). Further," he complains, "there is no agreed upon or satisfactory method for determining subtypes (McKinney, 1984)" (p. 197).

The demand, therefore, appears to be for more specific diagnosis and more specific remediation. These goals were characteristic of the discrete-point language theory of the 1960s in second and foreign language testing. Swanson (1988) shows that this same sort of thinking is current in the study of learning disabilities when he says, "Simply stated, a learning disability reflects a cognitive deficit...that is *reasonably* specific to a particular domain (e.g., reading).[3] The specific deficits displayed by such children must not extend too far into other domains of cognitive functioning. If they did, the concept of a learning disability would be meaningless..." (p. 196; his italics). However, Swanson goes on to observe that in fact "the literature has undermined the concept of specificity" (p. 197).

If we accept the major premise of Swanson that "the LD field is directed by social consensus" (p. 196), then it would follow that "the literature" which both establishes and defines the "consensus" could perhaps happily be redirected. However, I believe that it is not the "literature" per se that has "undermined the concept of specificity" as if there had been an active conspiracy against the "social consensus" that defines "the field of learning disabilities" (all the quoted terms being from Swanson, 1988). The evidence is simply against the idea of specificity in the way that it has been put forward. As argued extensively above, a more comprehensive and integrated view of semiotic capacities is needed to incorporate and explain rather than deny or purge the data of existing research.

A pragmatic approach, along the lines described above will be required, and the goal of isolating highly specific elements of cognition will generally have to be abandoned as a logical mistake. Cognition by its very nature involves the differentiation of specific elements only in rich and dynamic tensional contexts in which those elements find their distinctive identities. Apart from such contexts, those specific elements do not exist. This has been the primary motivation for clinical discourse analysis (Damico, 1985a, 1985b), an approach which seeks to understand the actual dynamics of the communicative performances of children rather than to pigeon-hole them into ready-made categories that may turn out to be altogether inappropriate in many cases. Discrete elements of cognitive processing only attain the character that really defines them in the contexts of their dynamic tensional oppositions in relation to each other and the whole continuum of experience (see the voluminous writings of Pierce on this matter as represented in collections by Burks, 1958; Hartshorne and Weiss, 1931-1935; Fisch, et al., 1982; Moore, et al., 1984; and Oller, 1989).

What about the current consensus that defines and purports to identify children with language disorders and/or learning disabilities? While the latter category has come more by tradition than by evidence to be associated with "neurological impairment", the idea that the former category is a subset of the latter is merely a matter of definition. The distinction between the larger category, learning disabilities, and the subcategory, language disorders (cf. Rueda and Mercer, 1985; also Cummins, 1986, p. 29), is merely assumed to be generally valid.[4] The distinction is never demonstrated by factual evidence anywhere in the vast literature on learning disabilities. In addition to a critical examination of this distinction, therefore, I wonder about the social consensus that sustains (Swanson, 1988) the whole field of special education and the study of exceptionalities in general.

As soon as the National Advisory Committee on Handicapped Children (1968) launched the first sentence of its long-standing definition of "learning disabilities" the confounding of that term with "language proficiency" and therefore with "language disorders" should have been abundantly apparent. From there forward, the problem of providing a theoretically adequate basis for the sought after distinctions only becomes more confused. They wrote:

> Children with learning disabilities exhibit a disorder in one or more of the basic psychological processes involved in understanding or using spoken or written languages. These may be manifested in disorders of listening, thinking, talking, reading, writing, spelling, or arithmetic. They include conditions which have been referred to as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, developmental aphasia, etc. They do not include learning problems which are primarily due to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance, or to environmental disadvantage (p. 4).

What is remarkable is that a vast number of workers could be encouraged to entertain the illusion that the kind of thinking expressed by the NACHC (and similar bodies) was a sufficient foundation on which to erect the present superstructure of the vast and growing edifice of special education.

Coles (1978) reviewed ten of the most widely used procedures for identifying children with the sorts of "disabilities/disorders" supposedly defined in the previous paragraph. He examined the *Illinois Test of Psycholinguistic Abilities*, *Bender Visual-Motor Gestalt Test*, *Frostig Developmental Test of Visual Perception*, *Wepman Auditory Discrimination Test*, *Lincoln-Oseretsky Motor Development Scale*, *Graham-Kendall Memory for Designs Test*, *Purdue Perceptual Motor Survey*, *Wechsler Intelligence Scale for Children—Revised*, neurological evaluations, and electro-encephalograms. These were found to be the most common procedures in use for the identification and diagnosis of learning disabilities in most states.

The sad conclusion was that "the predominant finding in the literature suggests that each test fails to correlate with a diagnosis of learning disabilities" (p. 326). Neither was there evidence of correct diagnosis in the results of therapeutic interventions: "In experiments where the dysfunction itself was treated, there was little success" (p. 326). While correlation alone is never proof of a causal relation, the absence of correlation is fatal to theories about specific causal connections. At the end of his article, Coles asserted, somewhat optimistically it would seem in retrospect, that "there is little question that eventually the tests reviewed here will be discarded; the evidence against them is mounting" (p. 335). If we think in terms of centuries rather than decades, this statement may yet turn out to be correct. At the moment, the tests in question are probably being used in about as many states and in far more cases in 1991 than they were in 1978.

When it comes to the subset of learning disabilities known as language disorders, there is even more confusion, if that is possible. The deep underlying question is what do tests used to define language disorders (and learning disabilities) really measure? The theory is that they should measure something over and above whatever intelligence tests measure. According to most researchers they are supposed to identify actual "neurological impairments" or at least "neurological inefficiencies" (Swanson, 1988).

However, if we take a paradigm exemplary test such as the *Illinois Test of Psycholinguistic Abilities* (Kirk, McCarthy, and Kirk, 1968), it turns out to be notably ineffective in predicting even reading scores if we control for IQ. Newcomer and Hammill (1975) reported that the correlation between ITPA scores and reading scores evaporated when intelligence was used as a covariate. Our point here is not to defend IQ tests as such (on the contrary, see part B below), but to show how confounded the constructs of language disorders, learning disabilities, and IQ are with each other. Moreover, we are arguing that all of these constructs have tended to overlook what is probably the single most important mediating variable, namely, primary language proficiency.

In general there has been a consensual distinction between "mental retardation" and "minimal brain damage" or "neurological impairment." Mental retardation is supposed to be related to, among other things, scores below some arbitrarily established level on standardized IQ scales. We, like Cummins (1984, see note 9, p. 30), do not deny that brain damage occurs in some cases or that mental retardation is in some instances a useful designation. What we do question, on the other hand, is whether these categories can be and are adequately distinguished on the basis of the present approach to IQ measurement and learning disabilities diagnosis (also see Mercer, 1973; Briere, 1973). There is substantial evidence that the distinction is thoroughly confounded in large numbers of cases. For instance, children identified as having "learning disabilities" in many cases are well below average in IQ scores. Out of 3,000 "learning disabled" children (identified as such in twenty-one states), more than a third fell below 90 on the standard IQ scale (Kirk and Elkins, 1975).

Why would educators tend to place at least some "mentally retarded" cases in the "learning disabled" category? It is clear that the former category is more stigmatized than the latter, and that the compassionate diagnostician, psychologist, or whatever, will prefer the less damaging label. But the problem surely runs much deeper than this. Beers and Beers (1980) point out that in some school systems a fourth to a third of the total school kindergarten population is being flagged as "potentially" learning disabled. This seems odd when a dramatically smaller percentage of the population is apt to have either genetic or acquired physical disabilities. Cummins (1984) aptly describes the category of "learning disabled," therefore as "a dumping ground for a wide variety of learning and behavioral difficulties" (see also Hallahan and Cruickshank, 1973). Swanson (1988) confesses that there is not a single trait, nor even a cluster of them, that can be identified as common to the category.

Undoubtedly it was because of the profound degree of confusion about the relation between mental retardation and learning disabilities that the American Association of Mental Deficiency arbitrarily changed the definition of "mentally retarded" from one to two standard deviations below the mean on a standardized IQ scale (McKnight, 1982). Cummins (1984, p. 83) sees this change as motivated by the desire to reclassify large numbers of formerly "mentally retarded" children as "learning disabled." A question that immediately arises is what such a change means in reference to the underlying constructs of intelligence versus neurological impairments. Beyond this, there is the lingering question of how language proficiency may be construed as relating to either of these constructs. What is disturbing is that in their educational applications both constructs are becoming, it would seem, increasingly folkloric and arbitrary.

Traditionally the identification of children with "language disorders" or "communicative disorders" or the general run-of-the-mill class of "learning disabilities" has been based on fairly superficial, surface-oriented criteria. For example, traditional diagnosticians have asked whether or not a child appropriately uses plural nouns (e.g., "dogs" versus "dog"), possessives (e.g., "Jim's hat" versus "Jim hat"), third person singular non-past verbs (e.g., "he walks" versus "he walk"), past tense verbs (e.g., "wanted" versus "want"), noun-verb agreement (e.g., "I am" versus "I be" or "I is"), irregular verbs (e.g., "fell" versus "falled"), number concord (e.g., "these cats" versus "this cats" or "these cat"), auxiliaries (e.g., "they have gone" versus "they gone," "they be gone," or "they done gone"). With respect to phonology, clinicians have tended to emphasize such things as the various forms of the regular plural morpheme in English (viz., /-z/, /-s/, or /-^z/) and the similar variations that occur in possessive marking of nouns, the third person singular non-past marking of verbs, the contractions of "is" and "has," and the similar variations that occur in marking of regular past-tense verbs (viz., /-d/, /-t/, or /-^d/).

Of course, surface form has some significance in its own right, but it has been elevated in the traditional tests, measurements, and diagnostic procedures of speech-language pathologists to such a position of prominence that the deeper purposes, the pragmatic aims of communication have been overlooked. As a result, "language disorders" have typically been defined in terms of superficial elements of syntax, morphology, and phonology, and more often than not have been strictly limited to problems of speech and writing rather than deeper aspects of the production and comprehension of meaningful discourse. Not only has the diagnostic definition of "language disorders" qua "learning disabilities" been based on surface-oriented criteria traditionally, but the treatment of them has likewise focussed on "intensive instruction in phonics" and "perceptual training" (cf. Beers and Beers, 1980, p. 73). The remedies, like the diagnoses, have been largely ineffective (Coles, 1978).

When attention is turned to discourse processing and to pragmatic criteria that have the potential at least of tapping into the deeper conceptual processes that underlie it, it is expected that the identification of genuine communicative difficulties, the kind that are apt to influence academic achievement in dramatic ways are more apt to be turned up (Damico and Oller, 1980; Damico, Oller, and Storey, 1983; Damico, 1985a, 1985b; Damico and Oller, 1986; McCord and Haynes, 1988[5]). This is not to say that researchers are presently in a position to determine on the basis of any existing testing program the specific neurological correlates of a given performance. This may be possible in rare cases but is certainly not the norm. Rather, as Coles (1978) intimated, there are no fully developed "less well-known instruments standing in the wings" (p. 335) and ready to fill the present void of thoroughly validated diagnostic procedures. As Coles said, "These tests, in any case, do not yet exist" (p. 335), and even the theory for their development is largely lacking.

What chiefly stands in the way of the needed theoretical and practical development is the uncritical acceptance of the present "social consensus." If researchers and practitioners alike are willing to acquiesce to the status quo of existing categories such as "language disorders," "learning disabilities," "mental retardation," and in general to the whole "diagnosis/remediation" paradigm, the needed reform of theory and practice is bound to be delayed if it ever comes at all. As Cazden (1985) has argued, the labeling of minority children especially as "disabled" or "disordered" must be, in her words, "delegitimized" and this can only be accomplished by looking to the broader context of socialization and education as has been argued by Coles (1978), Cummins (1984, 1986), and by Oller and Perkins (1978).

Based on all of the foregoing a few heuristic guidelines may be offered. Since the damage is likely only in cases of disabilities rather than giftedness, we concentrate on the former. To begin with there are logically just four types of errors to be avoided: (1) a LEP may be wrongly identified as disabled; (2) a truly disabled LEP child may be left out of the disabled category; (3) a LEP child may be incorrectly classed as a non-LEP; or (4) a non-LEP may be classed as a LEP.

It is known that large numbers of errors of type (1) are occurring. Many LEPs are incorrectly being diagnosed as disabled, or otherwise retarded. It follows from the same studies documenting type (1) errors that type (2), disabled LEPs not being identified as such, must also be common. Error type (3), LEPs incorrectly classed as non-LEPs, seems most likely when in Cummins' terms a child has developed substantial BICS in English but not much CALP. In these cases educators are apt to be fooled into thinking the child is ready for literacy in English when the child is still below threshold even in his or her primary language. Error type (4), non-LEPs classed as LEPs, can also occur if the child is evaluated on the basis of limited BICS while well-developed CALP in the child's primary language may be overlooked. The likelihood of a growing number of misclassifications of all four types is on the upswing due to the increasing number of non-English speaking minorities in our schools.[6]

To minimize errors of all four types a series of assessment phases is recommended. In all phases, the pursuit of evidence concerning the child should be treated in a matter-of-fact manner and with a view to the advocacy of the interests, needs, and feelings of the child above those of the school or the diagnostician. Our purpose as educators should be to promote and guard the interests of the child, not those of some abstract political or educational entity such as a state, institution, profession, or psychological yardstick (Cazden, 1975; Coles, 1978; Cummins, 1986).

First, to distinguish LEPs from non-LEPs, a variety of sources of evidence should be considered, e.g., talk with the child, observe the child's behavior in casual contexts, talk with siblings, parents, friends, etc., where appropriate. Ask about literacy and previous educational experience. Keep in mind that superficial, routine verbal skills may be deceptive in two ways: (a) they may lead us to attribute more language ability than is really present, or they may seem to indicate a low level of academic readiness when in fact the child is already literate in one or more other languages. Clear-cut cases may be decided on the basis of this preliminary phase to be either LEP or non-LEP. Doubtful cases should be referred to the second phase of assessment.

Two kinds of doubtful cases may be distinguished. Children with substantial educational background, e.g., those who have attained literacy in one or more other languages, but who lack basic routine skills (BICS) in English constitute the first case. These children should be evaluated with reference to their attainment in their most developed or primary language(s). For instance, some Asians will prove to be weak in English but literate in French and possibly some other language. To determine this fact may require additional interviews and possibly testing in the primary language. The question to be addressed in these cases is presumably, would it best serve the interests of this child if he or she were mainstreamed? If Cummins (1984 and elsewhere) is correct in the threshold hypothesis, only children who have demonstrated fairly advanced literacy skills or other abstract linguistic capabilities should be mainstreamed.

The other kind of doubtful cases referred from phase one would include the children who appear to have substantial ability to perform routine tasks in English (BICS) but who may or may not be ready for academic mainstreaming. The determination here, as in all cases, should be based on the solution that is believed most likely to benefit the child optimally. Preferences on the part of the child, and or the child's parents, should be weighed together with further evidence concerning academic readiness. The latter should be evaluated mainly in terms of the child's ability to perform abstract reasoning in the primary language and/or in English. Again, if Cummins (1984) is on the right track and if the theory as discussed above is followed in a general way, well-developed abstract reasoning capacities in one language will easily transfer to another assuming that there are no affective or social barriers[7] actively interfering with the process. In short, presumably some of these children should be mainstreamed, and some should not.

Phase three concerns children who have been identified as LEPs needing some kind of special program to enable them to profit optimally from their on-going educational experience. The objective during this phase is to differentiate children who are ready for a normal course of instruction in their primary language and those who may need some extra help beyond this. The latter are those traditionally labeled "learning disabled."

At this point, teachers or competent para-professionals who know the primary language(s) of the children should have already been involved and now become the main assessors. They should be trained in the deeper kinds of language assessment procedures that look to discourse/text-based tasks that include the broad range of communicative activities that school children are becoming able to engage in: e.g., relating an experience, singing a song, reading and reacting to a story, drawing a picture to illustrate some idea, explaining an illustration, evaluating a facial expression or gesture in a filmed narrative, play or drama, writing a letter, answering an advertisement, etc. The list of tested activities should be as broad as the curriculum children are expected to cope with. As suggested by Damico, Oller, and Storey (1983) and elaborated by Damico and Oller (1985) as well as Damico (1985a, 1985b, 1991) LEP students should be assessed in all of their languages and in each case across the broad spectrum of abilities so as to identify strengths. The objective at all points along the way should be not to look merely at surface forms but to look more deeply into the pragmatic aspects of discourse processing.

If there is even the slightest clue that the child is bilingual or multilingual every effort must be made to test the child in his or her strongest language(s). Some probing on this point may be necessary since it may not occur to the child, or to his parents, to tell some teacher or diagnostician, "By the way, I can read and write in Mandarin." They may not see this fact as relevant in an English speaking society or school. It may, however, be of considerable importance to an appropriate assessment of the child's actual capabilities. If a "disability" is suspected, where children are thoroughly bilingual or even multilingual, it is mandatory to assess their abilities in each of the languages they know. Usually this will involve only English and one other language, but in exceptional cases three or even more languages might be involved. To make a convincing case for a "learning disability," it is necessary to show that problems appearing in one of the child's languages also appear in the other.

There is no theory of language acquisition that will support the thesis that "learning disabilities" will only be manifested in French, or any other particular language. Deep semiotic processing problems, the kind that affect language capacity in a general way, or possibly other semiotic representational processes as well, are bound to manifest themselves in a variety of ways and cannot logically be limited to just one of a multilingual's languages. On the other hand, if problems are just apparent in one of two or more language systems a child possesses, it follows that the difficulties are likely to be within the normal range experienced by second language learners and that no real "learning disability" exists at all.

Phase four, for children identified as having special semiotic problems in more than one language or other semiotic modality, would involve a complete discourse analysis along the lines of Damico (1985a, 1985b, 1991) leading into recommendations for therapeutic intervention of an appropriate sort. At this point assessment merges with instruction (alias therapy) so completely that the two can no longer be profitably distinguished.

It would seem that procedures for intervention could benefit as much from an investigation of language instructional methods that work (cf. Oller and Richard-Amato, 1983; and Richard-Amato, 1988) as assessment of abilities and disabilities of LEPs could from the findings of language testing research. More particularly, pragmatically motivated procedures that deal with problems in the full richness and scope of normal experience will have a far better chance of success than discrete-point oriented procedures that are generally acknowledged to be recipes for failure (see Coles, 1978).

Here are a few heuristic guidelines for assessment in general. Samples of discourse, or assessment procedures themselves, should always involve performances in engaging contexts of semiotic representation. Wherever possible a variety of sources of evidence should be examined, e.g., multiple languages, dialects, kinesic representations, and sensory-motor performances. The objective should always be to find the child's optimal capabilities not to define some set of disabilities. Judgments should never be considered final but should be subject to constant updating, revision, and rechecking. No single test should form the basis for assessment. It should not be the basis for any final judgment. In the final analysis our goal is to set the child up for success, not for failure.

Notes

[1] Interestingly, Olson (1986) goes even further than Oller (1981). Subsequently, however, I believe we have followed the same river of thought (see Oller, 1989; Olson, 1986; Langer, 1987; and Sternberg, 1987).

[2] According to an unpublished study reported on at this meeting by Dr. Sherry R. Migdail, as few as 50 out of 1,000 students in a typical middle America school district were observed to have some form of genuine special education need (e.g., mental retardation, language-disorder/learning-disability, etc.). Yet, as Dr. Alba Ortiz observed in her presentation at this conference, a far higher percentage of students are misidentified as needing special education.

[3] Of course, the implication that a term like "reading" (or "listening" or even "spelling," all of which occur elsewhere in Swanson's paper) can be construed as "specific" is absurd on its face. Reading is as complex as any process known to modern science. Neither is it distinguishable except in superficial ways from all that accompanies it -- reasoning, arguing, imagining, etc. To suggest that such a process achieves the sought after "specificity" is to reveal the shallowness of thinking that characterizes the whole "social consensus" that constitutes "the field of LD".

[4] Cummins (1986) cites Rueda and Mercer (1985) who claimed that the distinction between "learning disabled" and "language disordered" for minority children is typically a matter of whether there is a "psychologist" or a "speech-pathologist" on the placement committee. Cummins concludes that the distinction is essentially arbitrary" (1986, p. 29).

[5] It should be noted that the latter authors, according to their own bibliography, only had access to summarial presentations of the pragmatic criteria they attempted to employ. Also, they compared only 12 "learning disabled" children as determined by the criteria set by the State of Alabama with 12 normals defined as such in view of their performance at "expected academic grade level". The authors apparently assume, without justification, that the children

identified by the state's criteria really are "learning disabled," but this is precisely the premise that needs to be questioned. Unless independent evidence of "learning disability" exists in those 12 children, evidence that would be missing for the "normals" against whom they are to be compared, the pragmatic criteria for evaluation cannot be tested with the experimental design that was in fact employed. In the final analysis, only some of the pragmatic criteria proposed by Damico and company did discriminate between the "disabled" and "normal" groups. However, this may be as much a consequence of group selection as of the criteria. Besides, it has been argued that significant difficulties can be expected for children that depart substantially from the norm on any one of the pragmatic criteria under consideration.

[6] Note that we do not use the term "disabled" here to legitimize it, nor do we agree that children in general to whom the label is attached are as it describes them. Our point here is to enable all children, LEPS and non-LEPs, normal and exceptional, to have access to the full range of educational benefits to which they are legitimately entitled.

[7] Krashen (1981, 1982, 1985) has argued that affective resistance to normal second language acquisition may occur in high anxiety or otherwise disturbing contexts. Assuming he is correct in this, every effort should be made to avoid the kinds of social conditions that might constitute or at least augment the mounting of such barriers.

References

Adams, C.. and Bishop, D.V.M. (1990). Conversational characteristics of children with semantic pragmatic disorder. I: Exchange structure, turntaking, repairs and cohesion. British Journal of Disorders of Communication, 24 (3), 211-239.

Alderson, J.C. (1983). The cloze procedure and proficiency in English as a second language. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 205-217). Rowley, MA: Newbury House.

Alderson, J.C. and Hughes, A., (Eds.), (1981). Issues in language testing. ELT Documents No. 111. London: British Council.

Anderson, J. (1969). Application of cloze procedure to English learned as a foreign language. Unpublished doctoral dissertation, University of New England, Australia.

Anderson, R.C. (1972). How to construct achievement tests to assess comprehension. Review of Educational Research, 42, 145-170.

Asher, J.J. (1969). The total physical response approach to second language teaching. Modern Language Journal 53, 3-17.

Asher, J.J. and Garcia, R. (1969). The optimal age to learn a foreign language. Modern Language Journal 53, 334-341.

Asher, J.J. and Price, B.S. (1967). The learning strategy of the total physical response: some age differences. Child Development 38, 1219-1227.

Audet, L.R., and Hummel, L.J. (1990). A framework for assessment and treatment of language learning disabled children with psychiatric disorders. Topics in Language Disorders, 10 (4), 57-74.

Auerbach, E.R. (1989). Toward a social-contextual approach to family literacy. Harvard Educational Review, 59, 165-181.

Bachman, L. (1990). Fundamental considerations in language testing. New York: Oxford.

Bachman, L. and Palmer, A.S. (1983). The construct validity of the FSI Oral Interview. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 154-169). Rowley, MA: Newbury House.

Bachman, L. and Clark, J.L.D. (1987). The measurement of foreign/second language proficiency. Annals of the American Academy of Political and Social Science 490, 20-33.

Baker, E.L. (1980). Introduction. In E.L. Baker and E.S. Quellmalz (Eds.), Educational testing and evaluation: Design, analysis, and policy. Beverly Hills, CA: Sage.

Baratz, J. (1969). A bidialectal task for determining language proficiency in economically disadvantaged Negor children. Child Development 40, 889-901.

Barrett, R. (1976). A new integrative test of writing ability. Paper presented at the Annual Meeting of Region V of the National Association of Foreign Students of America, Chicago.

Beers, C.S. and Beers, J.W. (1980). Early identification of learning disabilities: facts and fallacies. Elementary School Journal 81, 67-76.

Benbow, C.P., and Arjmand, O. (1990). Predictors of high academic achievement in mathematics and science by mathematically talented students: A longitudinal study. Journal of Educational Psychology, 82 (3), 430-441.

Binet, A. (1911). New investigation upon the measure of the intellectual level among school children. L'annee psychologique, 5, 145—201. Reprinted in The development of intelligence in children (274-329). Tr. E. Kite., Baltimore: Williams and Williams.

Binet, A. and Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveaux intellectuel des anormaux. L'anlaee psychologique 11, 191-24-i. Reprinted in The development of intelligence in children (41-88). Tr. E. Kite., Baltimore: Williams and Williams.

Bishop, D.V.M., and Adams, C. (1990). Conversational characteristics of children with semantic pragmatic disorder. II: What features lead to a judgment of inappropriacy. British Journal of Disorders of Communication, 24 (3), 211-239.

Bloom, A. (1987). The closing of the American mind. New York: Simon and Schuster.

Bloom, B.J. (1976). Human characteristics and school learning. New York: McGraw Hill.

Bloom, B.J. and Krathwohl, D.R. (1977). Taxonomy of educational objectives. New York: Longman.

Bormuth, J.R. (1970). On the theory of achievement test items. Chicago: University of Chicago Press.

Bower, T.G.R. (1971). The object in the world of the infant. Scientific American 225, 30-38.

Bower, T.G.R. (1974). Development in Infancy. San Francisco: Freeman.

Boyle, J.P. (1987). Intelligence, reasoning, and language proficiency. Modern Language Journal, 71, 277-88.

Briere, E. (1966). Quantity before quality in second language composition. Language Learning 16, 141-151.

Briere, E. (1973). Cross-cultural biases in language testing. In J.W. Oller, Jr. and J.C. Richards, (Eds.), (1973), Focus on the learner (pp. 214-227). Rowley, MA: Newbury House.

Brown, H.D. (1973). Affective variables in second language acquisition. Language Learning 23, (3) 1-244.

Bruner, J.S. (1984). Language, mind, and reading. In H. Goelman, A. Oberg, and F. Smith (Eds.), Awakening to literacy (pp. 195-201). Portsmouth, NH: Heinemann.

Burks, A.W., (Ed.), (1958). Collected writings of C.S. Peirce, Volumes VII and VIII. Cambridge, MA: Harvard University.

Carroll, J.B. (1961). Fundamental considerations in testing for English proficiency of foreign students. In Testing the English proficiency of foreign students (pp. 31-40). Washington, D.C.: Center for Applied Linguistics. Also in H.B. Allen and R.N. Campbell, (Eds.), Teaching English as a second language: A book of readings (pp. 313-320). New York: McGraw Hill.

Carroll, J.B. (1967). Foreign language proficiency levels attained by language majors near gráduation from college. Foreign Language Annals, 1, 131-151.

Carroll, J.B. (1983a). Psychometric theory and language testing. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 80-107). Rowley, MA: Newbury House.

Carroll, J.B. (1983b). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. Dillon and R. Schneck, (Eds.), Individual differences in cognition (pp. 1-28). New York: Academic.

Carroll, J.B., Carton, A.S., and Wilds, C.P. (1959). An investigation of cloze items in the measurement of foreign languages. College Entrance Examination Board Research and Development Report, Harvard University. ERIC ED 021513.

Carroll, J.B., and Sapon, S.M. (1959a). Modern Language Aptitude Test (Form A). New York: Psychological Corporation.

Carroll, J.B., and Sapon, S.M. (1959b). Modern Language Aptitude Test Manual . New York: Psychological Corporation.

Cazden, C.B. (1985). The ESL teacher as advocate. Paper presented at the Nineteenth Annual TESOL Convention, New York.

Chase, A. (1977). The legacy of Malthus: The social costs of the new scientific racism. New York: Knopf.

Chihara, T., T. Sakurai, and J.W. Oller, Jr. (1989). Background and culture as factors in EFL reading comprehension. Language Testing 6 (2), 143-151.

Chomsky, N. (1956). Three models for the description of language. In I.R.E. Transactions on Information Theory, vol. IT-2, pp. 113-124. Reprinted with corrections in R.D. Luce, R. Bush, and E. Galanter, (Eds.), (1965), Readings in mathematical psychology, vol. II, New York: Wiley.

Chomsky, N. (1957). Syntactic structures. The Hague: Mouton.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT.

Chomsky, N. (1972). Language and mind. New York: Harcourt, Brace, and Jovanovich.

Chomsky, N.A. (1975). Reflections on language. New York: Pantheon.

Chomsky, N. (1980a). On cognitive structures and their development. In M. Piatelli-Palmarini, (Ed.), Language and learning: The debate between Jean Piaget and Noam Chomsky (pp. 35-54). Cambridge, MA: Harvard.

Chomsky, N. (1980b). Rules and representations. New York: Columbia University.

Chomsky, N. (1988). Language and problems of knowledge. Cambridge, MA: MIT.

Church. A. (1951). The need for abstract entities in a semantic analysis. Daedalus &0, 100-112.

Clifford, Ray T. (1980). Language proficiency testing: a dynamic model. Paper presented at the Second International Language Testing Symposium, May 21-23, Darmstadt, Germany.

Cohen, A. (1980). Testing language ability in the classroom. Rowley, MA: Newbury House.

Coles, G.S. (1978). The learning disabilities test battery: Empirical and social issues. Harvard Educational Review 48, 313-340.

Condon, W.S. and Ogston, W.D. (1971). Speech and body-motion synchrony of the speaker-hearer. In D.L. Horton and J.J. Jenkins, (Eds.), The perception of language (pp. 150-173). Columbus, Ohio: Merrill.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of the research findings and explanatory hypotheses. Working Papers on Bilingualism No. 9, 1-43.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. Review of Educational Research 49, 222-251.

Cummins, J. (1981). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. Paper presented at the InterAmerican Language Proficiency Assessment Symposium, Arlie House, Virginia, March. In C. Rivera (Ed.), Communicative competence approaches to language proficiency assessment: Research and Application. Clevedon, England: Multilingual Matters.

Cummins, J. (1983a). Functional language proficiency in context: Classroom participation as an interactive process. In W. J. Tikunoff (Ed.), Compatibility of the SBIS features with other research on instruction for LEP students. San Francisco: Far West Laboratory.

Cummins, J. (1983b). Heritage language education: A literature review. Toronto, Canada: Ministry of Education.

Cummins, J. (1983c). Language proficiency and academic achievement. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 108-130). Rowley, MA: Newbury House.

Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy Clevedon, England: Multilingual Matters.

Cummins, J. (1986). Empowering minority students: A framework for intervention. Harvard Educational Review 56, 18-35.

Cummins, J. and Mulcahy, R. (1978). Orientation to language in Ukrainian-English bilingual children. Child Development 49, 1239-1242.

Cummins, J. and Swain, M. (1986). Bilingualism in education: Aspects of theory, research, and practice. London: Longman.

Davies, A. (1984). Review of J.W. Oller, Jr., (Ed.), Issues in language testing research (Newbury House, 1983). Language Testing 1, 111-114.

Damico, J.S. (1980, November). A clinical approach to discourse analysis in school age children. Mini-seminar presented at the American-Speech-Language-Hearing Association. Detroit, MI.

Damico, J.S. (1985a). Clinical discourse analysis: A functional approach to language assessment. In C.S. Simon (Ed.), Communication skill.s and classroom success (pp. 165-204). London: Taylor and Francis.

Damico, J.S. (1985b). The effectiveness of direct observation as a language assessment technique. Unpublished doctoral dissertation. University of New Mexico, Albuquerque.

Damico, J.S. (1991). Descriptive assessment of communicative ability in LEP students. In E.V. Hamayan and J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 157-218). Austin, TX: Pro-Ed.

Damico, J.S. and Oller, J.W. Jr. (1980). Pragmatic versus morphological/syntactic criteria for language referrals. Language, Speech, and Hearing Services in Schools 11, 85-94.

Damico, J.S. and Oller, J.W. Jr. (1985). Spotting language problems San Diego: Los Amigos Research Associates.

Damico, J.S., Oller, J.W., Jr., and Storey, M.E. (1983). The diagnosis of language disorders in bilingual children: pragmatic and surface-oriented criteria. Journal of Speech and Hearing Disorders 48, 285-294.

Davidson, F. (1988). An exploratory modeling survey of the trait structures of some existing language test data sets. Unpublished doctoral dissertation. University of California. Los Angeles.

Davies, A.(1984)

Duncan, S. E. and DeAvila. E. A. (1979). Bilingualism and cognition: some recent findings. NABE Journal 4, 15-50.

Ebbinghaus, H. (1897). Uber eine neue Methode sur Prufung geistiger Fahigkeiten und ihre Anwendung bei Schulkindern, Zeitschrift fur angewandte Psychologie, 13, 401-459.

Eco, U., Santambrogio, M., and Viola, P. (Eds.), (1988). Meaning and mental representations. Bloomington, IN: Indiana University Press.

Einstein, A. (1941). The common language of science. In Out of my later years (pp. 111-113). Secaucus, NJ: Citadel; also in J.W. Oller, Jr., (Ed.), (1989), Language and experience classic pragmatism (pp. 61-65). Lanham, MD: University Press of America.

Einstein, A. (1944). Remarks on Bertrand Russell's theory of knowledge. In P.A. Schilpp, (Ed.), The philosophy of Bertrand Russell (pp. 277-291). New York: Tudor Library of Living Philosophers; also in J.W. Oller, Jr., (Ed.), (1989), Language and experience: classic pragmatism (pp. 21-29). Lanham, MD: University Press of America.

ETS (1970). Manual for Peace Corps language testers. Princeton, New Jersey: Educational Testing Service.

Evola, J., E. Mamer, and B. Lentz. (1980). Discrete point versus global scoring for cohesive devices. In J.W. Oller, Jr. and Kyle Perkins (Eds.), Research in Language Testing (pp. 177-181). Rowley, MA: Newbury House.

Farhady, H. (1983a). The disjunctive fallacy between discrete-point tests and integrative tests. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 311-322). Rowley, MA: Newbury House.

Farhady, H. (1983b). New directions for ESL proficiency testing. In J. W. Oller, Jr, (Ed.), Issues in language testing research (pp. 253-269). Rowley, MA: Newbury House.

Farhady, H. (1983c). On the plausibility of the unitary language proficiency factor. In J W Oller, Jr., (Ed.), Issues in language testing research (pp. 11-28). Rowley, MA: Newbury House.

Finocchiaro, M. (1964). Teaching English as a second language. New York: Regents.

Fisch, M., et al., (Eds.), (1982). Writings of C.S. Peirce: A chronological edition. Volume 1. Indianapolis: Indiana University Press.

Fouly, K.A, Bachman, L.F., and Cziko, G.A. (1990). The divisibility of language competence: A confirmatory approach. Language Learning, 40 (1), 1-21.

Freire, P., and Macedo, D. (1987). Literacy: Reading the word and the world. S. Hadley, MA: Bergin and Garvey.

Galton, F. (1869). Hereditary genius. New York: Fontana Library.

Gardner, H. (1983). Frames of mind. New York: Basic Books.

Gardner, H. (1989). Assessment in context: The alternative to standardized testing. In B. Gifford (Ed.), Report to the Commission on Testing and Public Policy (cited by Gardner 1990 as 'forthcoming).

Gardner, H. (1990). The difficulties of school: Probable causes, possible cures. Daedalus, 119 (2), 85-113.

Gardner, H., and Hatch, T. (1989). Educational Researcher, 18 (8), 4-10.

Goodman, J.F., Freed, B., and McManus, W.J. (1990). Determining exemptions from foreign language requirements: Use of the Modern Language Aptitude Test Contemporary Educational Psychology, 15 (2), 131-141.

Goodman, K. (1967). Reading: A psycholinguistic guessing game. ,Journal of the Reading Specialist 4, 126-135.

Goodman, K. (1986). What's whole in whole language? Portsmouth, NH: Heinemann.

Goodman, K., and Goodman, Y. (1977). Learning about psycholinguistic processes by analyzing oral reading. Harvard Educational Review 47, 317-333.

Goodman, K., Goodman, Y., and Flores, B. (1979). Reading in the bilingual classroom: Literacy and biliteracy. Rosslyn, Virginia: National Clearinghouse for Bilingual Education.

Gould, S.J. (1981). The mismeasure of man. New York: W.W. Norton.

Gregg, K (1988). Epistemology without knowledge: Schwartz on Chomsky, Fodor, and Krashen. Second Language Research 4, 66-, 6.

Guilford, J.P. (1967). The nature of human intelligence. New York: McGraw Hill.

Gunnarsson, B. (1978). A look at the content similarities of between intelligence, achievement, personality and language tests. In J.W. Oller, Jr. and K Perkins, (Eds.), Language in education: testing the tests (pp. 17-35). Rowley, MA: Newbury House.

Guttman, L. (1965). The structure of interrelations among intelligence tests. In Educational Testing Service, Proceedings of the 1964 Invitational Conference on Testing Problems (pp. 25-36). Princeton, NJ: Educational Testing Service.

Hakuta, K. (1986). Mirror of language: The debate on bilingualism. New York: Basic Books.

Hakuta, K. and Diaz, R. (1984). The relationship between bilingualism and cognitive ability: a critical discussion and some new longitudinal data. In K. Nelson, (Ed.), Children's language, Vol. 5, Hillsdale, New Jersey: Lawrence Erlbaum.

Hallahan, D.P. and Cruickshank, W. (1973). Psychoeducational foundations of learning disabilities. Englewood Cliffs, New Jersey: Prentice-Hall.

Hamayan, E.V. and J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students. Austin, TX: Pro-Ed.

Hamp-Lyons, L. (1985). Two approaches to teaching reading: A classroom-based study. Reading in a Foreign Language, 3, 363-373.

Harris, D.P. (1985). Some forerunners of cloze procedure. Modern Language Journal 69 (4), 367-376.

Hartshorne, C. and Weiss, P., (Eds.), (1931-1935). Collected writings of C.S. Peirce, Volumes I-VI. Cambridge, MA: Harvard University.

Herrnstein, R.J. (1973). IQ in the meritocracy. Boston: Little Brown.

Herrnstein, R.J. and Wagner, A.R. (1981). Quantitative analysis of behavior. Cambridge, MA: Ballinger.

Hirsch, E.D. (1987). Cultural literacy. Boston: Houghton-Mimin.

Howey, KR. and Zimpher, N.L. (1989). Profiles of preservice teacher education: Inquiry into the nature of programs. Albany, NY: State University of New York.

Hughes, A. and Porter, D., (Eds.), (1983). Current developments in language testing (63-74). London: Academic.

Jakobson, R. (1980). The framework of language. Ann Arbor: University of Michigan.

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? Harvard Educational Review, 39, 1-23.

Jensen, A.R. (1980). Bias in mental testing. New York: Free Press.

John-Steiner, V. (1985). Notebooks of the mind: exploration of thinking. Albuquerque, NM: University of New Mexico Press.

Kagan, D.M. (1990). Ways of evaluating teacher cognition: Inferences concerning the Goldilocks Principle. Review of Educational Research, 60 (3), 419-469.

Kaufman, A S. (1979). Intelligent testing with the WISC-R. New York: Wiley.

Kessler, C. and Quinn, M.E. (1980). Positive effects of bilingualism on science problem-solving abilities. In J.E. Alatis, (Ed.), Georgetown University Roundtable on Languages and Linguistics. Washington, D.C.: Georgetown University.

Kirk, S.A and Elkins, J. (1975). Characteristics of children enrolled in the Child Service Demonstration Centers. Journal of Learning Disabilities 8, 630-637.

Kirk, S.A., McCarthy, J.J., and Kirk, W.D. (1968). Illinois Test of Psycholinguistic Abilities. Urbana: University of Illinois.

Kolers, P.A. (1968). Bilingualism and information processing. Scientific American 21, 78-84.

Krashen, S.D. (1981). Second language acquisition and second language learning. Oxford: Pergamon.

Krashen, S.D. (1982). Principles and practice in second language acquisition. Oxford: Pergamon.

Krashen, S.D. (1985). The input hypothesis: issues and implications. London and New York: Longman.

Kunnan, A.J. (1990). DIF in native language and gender groups in an ESL placement test. TESOL Quarterly, 24 (4), 741-746.

Lado, R. (1961). Language testing. New York: McGraw Hill.

Lambert, W. E. (1975). Culture and language as factors in learning and education. In A. Wolfgang, (Ed.), Education of immigrant students. Toronto: Ontario Institute for Studies in Education.

Lane, H. (1984). When the mind hears. New York: Random House.

Lane, H. (1988). Educating the American Sign Language Minority of the United States: A paper prepared for the Commission on the Education of the Deaf. In S. Wilcox, (Ed.), Academic acceptance of American Sign Language: Sign Language Studies, Special Issue 59 (pp. 221-230). Silver Spring, Maryland: Linstok.

Langer, J.A. (1987). A sociocognitive perspective on literacy. In J.A. Langer (Ed.), Language, literacy, and culture: Issues of society and schooling. Norwood, NJ: Ablex.

Lotz, J. (1951). Natural and scientific languages. Daedalus 80, 87-88.

Lowe, Pardee. (1980). Language proficiency testing: a static model. Paper presented at the Second International Language Testing Symposium, May 21-23, Darmstadt, Germany.

Luria, A.R. (1959). The directive function of speech in development and dissolution. Word, 15, 341-352.

Luria, A.R. (1961). The role of speech in the regulation of normal and abnormal behavior. New York: Irvington.

Luria, A.R. (1979). Verbal regulation of behavior. In M. Cole and S. Cole (Eds.), The making of a mind: A personal account of Soviet psychology (pp. 104-119). Cambridge, MA: Harvard University Press.

Luria, A.R. and Yudovich, F.A. (1959). Speech development and the mental processes of the child. London: Staples.

Lynch, B., Davidson, F., and Henning, G. (1988). Person dimensionality in language test validation. Language Testing, 5(2), 206-219.

MacNamara, J. (1973). The cognitive strategies of language learning. In J.W. Oller, Jr. and J.C. Richards, (Eds.), (1973), Focus on the learner (pp. 57-65). Rowley, MA: Newbury House.

MacNamara, J. (1982). Names for things: A study of human learning. Cambridge, MA: MIT.

MacNamara, J. (1989). Review of Meaning and mental representations (see Eco, Santambrogio, and Viola, 1988).pp. 237. ISBN 0-253-33724-0. Journal of Language and Social Psychology, 8(5), 349-353.

McCord, J.S. and Haynes, W.O. (1988). Discourse errors in students with learning disabilities and their normally achieving peers. Journal of Learning Disabilities. 21, 237-243.

McKinney, J.D. (1984). The search for subtypes of specific learning disabilities. Journal of Learning Disabilities 17, 43-51.

McKnight, R.T. (1982). The learning disability myth in American education. Journal of Education 164, 351-359.

Mercer, J.R. (1973). Labelling the mentally retarded. Los Angeles: University of California Press.

Mercer, J.R. (1984). What is a racially and culturally non-discriminatory test? A socialistic and pluralistic perspective. In C.R. Reynolds and R.T. Brown (Eds.), Perspectives on bias in mental testing (pp. 293-356). New York: Plenum.

Miller, L. (1990). The roles of language and learning in the development of literacy. Topics on Language Disorders, 10(2), 1-24.

Moore, E.C., et al., (Eds.), (1984). Writings of C.S. Peirce: A chronological edition. Volume 2. Indianapolis: Indiana University.

Nagel, E. (1959). Charles Sanders Peirce: A prodigious but little known American philosopher. Scientific American 200, 185- 192.

Nagel, E. (1979). Teleology revisited and other essays in the philosophy and history of science. New York: Columbia University.

National Advisory Committee on Handicapped Children. (1968). Special education for handicapped children (First Annual Report). Washington, D.C.: U.S. Department of Health, Education, and Welfare.

Newcomer, P.L. and Hammill, D.D. (1975). ITPA and academic achievement: a survey. Reading Teacher 28, 731-741.

Newmark, L. (1966). How not to interfere with language learning. International Journal of American Linguistics 32. Also in E.W. Najam and C.T Hodge. (Eds.), Language learning: The individual and the process. Bloomington, Indiana: Indiana University and in J.W. Oller, Jr. and P. Richard-Amato, (Eds.), Methods that work: A smorgasbord of ideas for language teachers (pp. 49-58). Rowley, MA: Newbury House.

Newmeyer, F. (1980). Linguistic theory in America. New York: Academic.

Nolen, S.B., and Haladyna, T.M. (1990). Personal and environmental influences on students beliefs about effective study strategies. Contemporary Educational Psychology, 15(2), 116-130.

Oller, J.W., Jr. (1970). Dictation as a device for testing foreign language proficiency. Workpapers in TESL: UCLA 4, 37-42. Also in English Language Teaching 25 (1971), 254-259.

Oller, J.W., Jr. (1972). Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. Modern Language Journal 56, 151-158.

Oller, J.W., Jr. (1973). Cloze tests of second language proficiency and what they measure. Language Learning 23, 105-118.

Oller, J.W., Jr. (1979). Language tests at school. London: Longman.

Oller, J.W., Jr. (1981). Language as intelligence? Language Learning 31, 465-492.

Oller, J.W., Jr. (1983a). A consensus for the 80's? In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 351-356). Rowley, MA: Newbury House.

Oller, J.W. Jr., (Ed.) (1983b). Issues in language testing research. Rowley, MA: Newbury House.

Oller, J.W., Jr. (1986). Communication theory and testing: What and how? In C. Stansfield, (Ed.), Proceedings of the second international invitational Conference on the Test of English as a Foreign Language (pp. 99-179).Princeton, New Jersey: Educational Testing Service.

Oller, J.W., Jr. (1988). The input hypothesis: Issues and implications by S.D. Krashen (1985). London and New York: Longman. Language 64, 171-173.

Oller, J.W., Jr. (Ed.), (1989). Language and Experience: Classic Pragmatism. Lanham, MD: University Press of America.

Oller, J.W., Jr. (1991). Language and bilingualism: More tests of tests. Cranbury, NJ: Bucknell University Press.

Oller, J.W.,Jr. and Conrad, C. (1971). The cloze procedure and ESL proficiency. Language Learning 21, 183-196.

Oller, J.W., Jr. and Perkins, K, (Eds.), (1978). Language in Education: Testing the Tests. Rowley, MA: Newbury House.

Oller, J.W., Jr. and Perkins, K, (Eds.), (1980). Research in Language Testing. Rowley, MA: Newbury House.

Oller, J.W., Jr., and Richard-Amato, P. (1983). Methods that work: A smorgasbord of ideas for language teachers. Rowley, MA: Newbury House.

Olson, D.R. (1986). Intelligence and literacy: The relationships between intelligence and the technologies of representation and communication. In R. Sternberg and R. Wagner (Eds.), Practical Intelligence. New York: Cambridge University.

Oltman, P.K, and Stricker, LJ. (1990). Developing homogeneous TOEFL scales by multidimensional scaling. Language Testing, 7(1), 1-12.

Oltman, P.K, Stricker, L.J., and Barrows, T.S. (1990). Analyzing test structure by multidimensional scaling. Journal of Applied Psychology, 75, 21-27.

Ortiz, A.A. and Yates, J.R. (1983). Incidence of exceptionality among hispanics: Implications for manpower planning. NABE Journal 7, 41-54.

Palmer, A.S. Groot, P.J.M. and Trosper, G.A., (Eds.).(1981). The validation of oral proficiency tests: Selected papers from The Colloquium on the Validation of Oral Proficiency Tests. Washington, D. C.: TESOL.

Peirce, C.S. (1878). How to make our ideas clear. Popular Science, Monthly, 12, 286-392. In Hartshorne and Weiss (1934), 248-271 and in Oller (1989), 53-58, 99-106.

Peirce, C.S. (1906). Truth. In Hartshorne and Weiss (1931), 3-18 and in Olier (1989), 193-201..

Piaget, Jean. (1947). The psychology of intelligence. Tottowa, New Jersey: Littlefield Adams.

Piatelli-Palmarini, M. (1980). Language and learning: The debate between Jean Piaget and Noam Chomsky. Cambridge, MA: Harvard University.

Politzer, R.L., Hoover, M.R., and Brown, D. (1974). A test in proficiency in black standard and non-standard speech. TESOL Quarterly, 8, 27-35.

Porter, D. (1983). The effect of quantity of context on the ability to make linguistic predictions: a flaw in a measure of general proficiency. In A. Hughes and D. Porter, (Eds.), Current developments in language testing (pp. 63-74). London: Academic.

Prestine, N.A. (1990). The struggle for control of teacher education: A case study. Educational Evaluation and Policy Analysis, 11(3), 285-300.

Prutting, C.A., and Kirchner, D.M. (1987). A clinical appraisal of the pragmatic aspects of language. Journal of Speech and Hearing Disorders, 52, 105-119.

Purcell, E.T. (1983). Models of pronunciation accuracy. In J. W. Oller, Jr., (Ed.), Issues in language testing research (pp. 133-153). Rowley, MA: Newbury House.

Quintero, E., and Huerta-Macias, A. (1990). All in the family: Bilingualism and biliteracy. The Reading Teacher, 44(4), 306-312.

Rand, E. (1972). Integrative and discrete-point tests at UCLA. Workpapers in TESL: UCLA 6, 67-78.

Rand, E. (1976). A factored homogeneous items approach to the UCLA ESL Placement Examination. Paper presented at the Tenth Annual TESOL Convention, New York City.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press.

Richard-Amato, P. (1988). Making it happen. New York: Longman.

Rist, Ray C. (1982). On the application of ethnographic inquiry to education: procedures and possibilities. Journal of Research in Science Teaching 19(6), 439-450.

Roid, G. H., and Haladyna, T.M. (1982). A technology for test item writing. New York: Academic Press.

Rueda, R. and Mercer, J.R. (1985). Predictive analysis of decision-making with language minority handicapped children. Paper presented at the BUENO Center Third Annual Symposium on Bilingual Education, Denver.

Russell, B. (1919). Introduction to mathematical logic. London: Allen and Unwin.

Sang, F., Schmitz, B., Vollmer, H.J., Baumert, J., and Roeder, P. M. (1986). Models of second language competence: a structural equation. Language Testing 3, 54-79.

Savignon, S. (1971). A study of the effect of training in communicative skills as part of a beginning college French course on student attitude and achievement in linguistic and communicative competence. Unpublished doctoral dissertation, University of Illinois, Champaign-Urbana, Illinois.

Savignon, S. (1983). Communicative competence: Theory and classroom practice (Texts and contexts in second language teaching). Reading, MA: Addison-Wesley.

Schumann, J. (1975). Affective factors and the problem of age in second language acquisition. Language Learning 26, 135-143.

Scott, R. (1979). Measures of Adult English Proficiency: Final report. Washington, D C.: Office of Education.

Scovel, T. (1988). A time to speak: A psycholinguistic inquiry into the critical period for human speech. Cambridge, MA: Newbury House.

Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics 10, 209-231.

Shepard, L.A. and Smith, M.L. (1983). An evaluation of the identification of learning disabled students in Colorado. Learning Disability Quarterly 6, 115-127.

Shuy, R.W. (1978). Problems in assessing language ability in bilingual education. in H. Lafontaine, H. Persky, and L. Golubchick, (Eds.), Bilingual education Wayne, New Jersey: Avery.

Shuy, R.W. (1981). Conditions affecting language learning and maintenance among Hispanics in the United States. NABE Journal 6, 1-18.

Simich-Dudgeon, C. (1987). Involving LEP parents as tutors in their children's education. ERIC/CLL News Bulletin 10 (2), 3-4.

Skutnabb-Kangas, T. and Toukomaa, T. (1976). Teaching migrant children's mother tongue and learning the language of the host country in the context of the socio-cultural situation of the migrant family. Helsinki: The Finnish National Commission for UNESCO.

Smith. F (1975). Twelve easy ways to make learning to read difficult, and one difficult way to make it easy. In F. Smith, (Ed.), Psycholinguistics and reading. New York: Holt, Rinehart, and Winston.

Smith, F. (1978). Understanding Reading. Second edition, New York: Holt, Rinehart, and Winston.

Smith, F. (1982). Writing and the Writer. New York: Holt, Rinehart, and Winston.

Smith, F. (1984). The creative achievement of literacy. In H. Goelman, A. Oberg, and F. Smith (Eds.), Awakening to literacy (pp. 135-142). Portsmouth, NH: Heinemann.

Smith, F. (1989). Overselling literacy. Phi Delta Kappan, 70, 352-359.

Spearman, C. (1904). "General intelligence" objectively determined and measured. American Journal of Psychology 15, 201-293.

Spearman, C. (1927). The abilities of man. New York: Macmillan.

Spolsky, B. (1968a). Language tests: The problem of validation. TESOL Quarterly 2, 88-94.

Spolsky, B. (1968b). What does it mean to know a language? Or how do you get someone to perform his competence? Paper read at the Second Conference on Problems of Foreign Language Competence, University of Southern California, Los Angeles. Also in J.W. Oller, Jr. and J.C. Richards, (Eds.), (1973), Focus on the learner (pp. 164-176). Rowley, MA: Newbury House.

Spolsky, B. (1983). Fourteen years on—later thoughts on overall language proficiency. In A. Hughes and D. Porter, (Eds.), Current developments in language testing (pp. 39-42). London: Academic.

Spolsky, B. (1985). A critical review of Krashen. Lecture presented at the University of New Mexico.

Sternberg, R.J. (1987). Second game: A school's eye view of intelligence. In J.A. Langer (Ed.), Language, literacy, and culture: Issues of society and schooling. Norwood, NJ: Ablex.

Stump, T. A. (1978). Cloze and dictation tasks as predictors of intelligence and achievement. In J.W. Oller, Jr. and K. Perkins, (Eds.), Language in education: testing the tests (pp. 36-64). Rowley, MA: Newbury House.

Swain, M., Dumas, G., and Naiman, N. (1974). Alternatives to spontaneous speech: elicited translation and imitation as indicators of second language competence. Working Papers on Bilingualism 3, 68-79.

Swanson, H.L. (1988). Toward a metatheory of learning disabilities. Journal of Learning Disabilities 21, 196-209.

Taylor, W. (1953). "Cloze procedure": A new tool for measuring readability. Journalism Quarterly 30, 415-433.

Taylor, W. (1956). Recent developments in the use of"cloze" procedure. Journalism Quarterly 33, 42-48, 99.

Taylor, W. (1957). "Cloze" readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology 41, 19-26.

Terrnan, L.M. (1925). Genetic studies of genius. Palo Alto: Stanford University.

Terrnan, L.M. and Merrill, M.A. (1960). Stanford-Binet intelligence scale: manual for the third revision, form L-M. Boston: Houghton Mifflin.

Terman, L.M. and Oden, M.H. (1947). The gifted child grows up. Palo Alto: Stanford University.

Trabue, M.R. (1916). Completion test language scales. New York: Teachers College, Columbia University.

Thurstone, L.L. (1924). The nature of intelligence. London: Paul Trench and Trubner.

Thurstone, L.L. (1938). Primary mental abilities. Chicago: University of Chicago.

Thurstone, L.L. (1947). Multiple-factor analysis: A development and expansion of the vectors of mind. Chicago: University of Chicago.

Thurstone, L.L. and Thurstone, T.G. (1941). Factorial studies of intelligence Psychometric monographs, no. 2. Chicago: University of Chicago.

Upshur, J.A. (1967). English language tests and the prediction of academic success. In National Association of Foreign Student Affairs, Selected conference papers of the Association of Teachers of English as a Second Language.

Upshur, J.A (1969a). Measurement of oral communication. In H. Schrand, (Ed.) IFS documentation, Leistungtionzenstaum fur Fremdsprachenforschung (53-80). House.

Upshur, J.A. (1969b). Productive communication testing: progress report. Paper presented at the Second International Congress on Applied Linguistics, Cambridge, England. Also in J.W. Oller, Jr. and J.C. Richards, (Eds.), (1973), Focus on the learner (pp. 177-183). Rowley, MA: Newbury House.

Upshur, J.A. (1979). Functional proficiency theory and a research role for language tests. In E.J. Briere and F.B. Hinofotis, (Eds.), Concepts in language testing: Some recent studies (pp. 75-100). Washington, D. C.: TESOL.

Upshur, J.A. and Fata, J. (Eds.), (1968). Problems in foreign language testing (Special Issue Number 3). Language Learning. Ann Arbor: University of Michigan.

Upshur, J.A and Homburg, T.J. (1983). Some relations among language tests at successive ability levels. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 188-202). Rowley, MA: Newbury House.

Valette, R. (1964). The use of dictee in the French language classroom. Modern Language Journal 39, 431-434.

Valette, R. (1967). Modern Language Testing New York: Harcourt, Brace, Jovanovich.

Vigil, N. and Oller, J.W.,Jr. (1976). Rule fossilization: A tentative model. Language Learning 26, 281-295.

Vollmer, H.J. and Sang, F. (1983). Competing hypotheses about second language ability: a plea for caution. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 29-79). Rowley, MA: Newbury House.

Vygotsky, L.S. (1934). Thought and language. Moscow-Leningrad: Sozekgiz. Tr. by E. Hanfmann and G. Vakar and printed in English (1962), Cambridge, MA: MIT.

Vygotsky, L.S. (1978). Mind in society: The development of higher psychological processes. Edited by M. Cole, V. John-Steiner, S. Scribner, and E. Souberman. Cambridge, MA: Harvard University.

Walters, J., and Gardner, H. (1985). The development and education of intelligences. In F. Link (Ed.), Essays on the intellect (pp. 1-21). Washington, D.C.: Curriculum Development Associates.

Walters, J., and Gardner, H. (1986a). The crystalizing experience. In R.J. Sternberg and J.E. Davidson (Eds.), Conceptions of giftedness (pp. 306-331). New York: Cambridge, University Press.

Walters, J., and Gardner, H. (1986b). The theory of multiple intelligences: Some issues and answers. In R.J. Sternberg and R. Wagner (Eds.), Practical intelligence: Origins of competence in the everyday world (pp. 163-182). New York: Cambridge University Press.

Wells, G. (1986). The meaning makers. Portsmouth, NH: Heinemann.

Whipple, Guy Montrose. (1915). Manual of mental and physical tests. Part II. Complex processes. Second edition, revised. Baltimore, MD: Warwick and York.

Wilcox, S., (Ed.), (1988). Academic acceptance of American Sign Language: Sign Language Studies, Special Issue 59. Silver Spring, Maryland: Linstok.

Response to John Oller's Presentation

Fred Davidson
University of Illinois, Urbana/Champaign

Well, half of me wants to say sort of this is really easy because I agree. I do agree very deeply. The title of my talk is, "From the Trenches." In this paper for this meeting, John Oller has presented a thorough, theoretical and philosophical basis for motivated, proactive change in the assessment of language minority students in the United States. In my reaction, I shall do two things. I'm going to briefly summarize and interrupt his main points, give you a glimpse of the rest of that 66-page document, and then attempt to relate his philosophical stance to the pragmatic necessities of language minority students. Now, it is this second section that has caused me to title my paper, "From the Trenches." Oller's work is broad reaching and provocative. From my background as a language tester, who has worked with small and large language assessment data sets, I have decided to challenge myself and discuss how his proposal might be implemented in the front line trench battles of language testing in school setting.

First the summary: Oller's paper has three parts. In the first part, he reviews primary and non-primary language testing literature. He discusses the heritage which language testing shares with intelligence measures as well as the historical link of language testing with structural linguistics. These two trends are primarily responsible for the prevalence of discrete-point testing approaches, and the question John raises or implies many times is -- is it appropriate to consider language ability as the sum of many parts? By the second section of the paper, Oller's beliefs are clear, when on page 20, he says, "Happily a movement toward pragmatic, holistic testing is now discernable."

Much of the first section of his paper seems directed at this conclusion -- a conclusion which I share very deeply. Language ability is indeed a complex mental trait and holistic integrative testing should hold forth more than it does. Oller says near the end of his paper, "It is difficult to over estimate the pervasive influence of analytic, discrete-point thinking in the study of exceptionalities." And it is precisely this pervasive influence that I've taken as my mandate: how to expand the framework of language assessment measures in the reality of school based decision making, and I am going to return to this later.

Second, Oller has a review of relevant points from the recent history of educational measurement. He cites Roid and Haladyna, "There is a chance for endless mapping sentences, facts, and facet elements with lack of agreement among developers being a major determent to progress." And then he goes on to say," when the focus is shifted from a list of items, (which is a poor characterization in any case of any non-finite domain of sentences) to the generative basis which underlies the representations that constitute that domain, we have some hope of achieving reliability and validity." The multi-facet nature criterion reference measurement or, strictly speaking, a domain referencing, is anathema to good language testing, Oller seems to say. I agree generally with this, but I suggest that criteria can also be holistic, and I've done some work in the design and implementation criterion reference test specifications that are pragmatic and holistic. The other major component of this section is citation of the work of Gardner on multiple intelligences as that is central to part three. I want to deal with it in my discussion of that part.

In part three, Oller sketches his own model of human systems of representation, and there are three diagrams in there ending at the one that integrates Gardner with Oller. He calls this his own general semiotic capacity model. This is by far the most philosophically challenging section of the paper. My impression is that Oller is expanding on the notion of a general factor of language to encompass multiple components, and that he is utilizing Gardner's work to do so. Oller now views language as a global factor that contains components. This is very clear in the paper and this is very welcome. He closes with a series of assessment recommendations for teaching and testing language minority students. Much of this discussion centers around the nature of disabilities. He claims that the handling of language minority students has been heavily conditioned by the history of measurement, of language disorders, and/or learning disabilities. I have seen this, first hand, in my work with K through 12 ESL and bilingual data in the state of Illinois. I agree heartily.

He has several recommendations at the very end, including the use of "pragmatically motivated procedures" that deal with the problems in the full richness and scope of normal experience as well as a call for multiple measures about which I will speak specifically below. He seems to regret the difficulty of implementing change in language minority student education. The ease with which a disability or remediation paradigm can rule the day prompts him to say, "What chiefly stands in the way of the needed theoretical and practical development is the uncritical acceptance of the present social consensus, if researchers and practitioners alike are willing to acquest to the status quo of existing categories like language disorders, learning disabilities, mental retardation, and in general to the whole diagnosis and remediation paradigm the needed reform of theory and practice is bound to be delayed if it ever comes at all." He is challenging our field then to find a way to break the uncritical acceptance of the status quo, and I'd like to take up that challenge in part in the next section.

So, a voice from the trenches.

Now, the issue here, it seems to me, is that we need to get inside the head of the people that matter. All assessment is done within the context of decision making. There is a real good paper by Jack Upshur from 1970 and Lyle Bachman extends it in his 1990 textbook. The person making the decision may or may not be a test designer and, if so, may or may not subscribe to the philosophical shifts which Oller promotes, and with which I heartily agree, so this begs the question, why? What causes the acquiescence that bothers John Oller and bothers me? Let me offer a practical, real world answer. I believe that we need to legitimize the change necessary for the assessment of language minority students. This legitimization requires two components. First, full-proof argument and second, logistical ease, i.e., that the new must be as easy to implement as the old. First, full-proof argument should affect assessment score users on a philosophical strong ground as Oller has done as well as be an elegant simplicity, and I'd like to offer an example of the later. Drawing heavily upon an excellent paper in Language Testing by Mats Oscarson, 1989. I highly recommend it. Oscarson argues that if modern language teaching is more focused on the learner then the learner should be consulted in the assessment process. He argues, therefore, that language testing should include self-report. At the very beginning of his paper, Oscarson notes that there are fundamentally two types of assessment, external and internal. The former, external, is imposed from outside of the learner. Most tests are actually external. The latter are self-report of some sort or another and reflects the internal goals, agenda, and motivations of the learner, goals which may or may not match the external tests. Oscarson's paper closes with samples of self-report and language testing and the particular appropriacies of those samples to K through 12 is not really relevant here. What is at issue here is the undeniable simplicity of Oscarson's argument. The differentiation of assessment into self and non-self in my eye is equal to the philosophical paradigm shift that separated criterion referencing from norm referencing. Hudson and Lynch, Language Testing, 1984 and Glaser 1963, whom they cite, note the following about the difference between norms and criteria. They note that if achievement happens in the classroom then a normalizing test will actually unskew a curve. All teachers after they teach want people to achieve. Apply a normalizing norm referenced test to that, and you will actually convert it back to a Bell Curve. Affectively, the achievement will be statistically squashed. That's a powerful argument which appeals to teachers everywhere. I maintain that Mats Oscarson's argument, that testing needs to be internal and external, is equally simple and powerful.

Several years ago, I was fortunate to be in a seminar with John Oller at UCLA. There I presented a case for something I then called and still call "multiple referencing" which is a super-ordinate term to link criterion referencing and norm referencing and other references, as yet to be determined. I believe that self-report is actually a form of test reference, call it self-referencing, on equal stature to that of criteria and norms. Furthermore, I believe the simple elegance of Oscarson's argument elevates self-referencing to the status of norms and criteria. The simple, elegant, undeniable elevation of the new to the status to the old is one crucial component to breaking the acquiescence which Oller condemns. In this particular instance of proposed change, I believe multiple referencing is not really a new concept just a new term. Oller even appeals for it at the very end of his paper, as have many others who have used terms like multiple criteria and multiple indicators, and I have a whole scad of references here on that. I maintain that terms like multiple indicators and multiple criteria help us see multiple sources of evidence within a certain score reference, but why not attack the number of references as well, and that's why I proposed self-referencing.

But this isn't enough. An argument in favor of expansion of the number of score references, which in essence, John does at the end of his paper, is not the only necessity by far. We need to make the change work. I often pose the following question to my language testing students. Two situations; Situation A: You are an administrator at a school, a decision maker. You have 900 new international students arrive at your school, and you must decide their English proficiency. You consult a single norm referenced discrete-point test score. Situation B: You are the same person. You have 900 new international students arrive at your school, and you must decide their English proficiency. You consult a single norm referenced test score, a single criterion referenced test score, and you interview each student for self-report. The issue is that the entire technological history of logistical ease and human measurement is intertwined with the summative discrete-point test score. We cannot get away from what we appear to do so well. Clinical, detached, quasi objective discrete-point norm referenced testing, we have that down pat. A couple of years ago at a conference, I met Edward DeAvila, a developer of the LAS assessment battery. He showed me a computer expert system program to help a decision maker navigate multiple information sources, some of which constituted multiple references. As I recall, he had both norms and criteria, and I think that this program was or was a refinement of one developed for the Chicago Public School System. Now, I'm not proposing, necessarily, that a computerized expert system can automate the navigation of multiple references and a broader range of what John calls pragmatically motivated procedures, but I do claim that unless we do something to break the logistical strangle-hold of norm referenced summative discrete-point tests, we are doomed to fail. Let's hit them with both barrels. Let's use the elegant simple arguments, and let's make routine the complexity of dealing with language testing as it should be dealt with.

In closing, I would like to echo the sentiment of Anne Frank, "I do believe that people are basically good at heart." ...despite the way this sounds. I do agree with Anne Frank. People are basically good at heart, and this includes the staunchest decision maker/addicts of norm referenced test scores. I believe, rather, that what happens is not that they consciously reject the persuasion of Oscarson, Oller, and others, but rather that such change is felt to be logistically impossible. Let's work on that feeling.

Response to John Oller's Presentation

Myriam Met
Montgomery County Public Schools, Maryland

I agree with both John Oller and Fred Davidson. I'm just a simple practitioner, so what I'd like to do with you this morning is try to extrapolate some of the implications from foreign language practice, from the paper that John has written, and from the remarks Fred has shared with you this morning. I'd like to talk briefly about the notion of global testing of proficiency and tie that to what I think is a more important and valuable trend for all of us, which is classroom assessment of language skills.

The first part, global proficiency, I think, draws from the buzz word in the foreign language profession today (and it has been for the last decade), which is "proficiency." You might find this definition interesting because it's a somewhat different view of the term "proficiency" from the one that I was familiar with when I worked in ESL and bilingual programs (about six years ago). In foreign language proficiency, one is never "proficient." One is only proficient to perform certain tasks or language functions, in certain contexts or settings about certain topics or contents, and with a degree of both linguistic and socio-cultural accuracy. To some extent, all of us are limited proficient in that none of us, even the most ideal, (but non-existent) educated native speaker is ever completely proficient to perform all language tasks, in all contexts, in all contents, with the same degree of linguistic and socio-cultural accuracy. That is an important concept to which I'll come back in a little while when I talk about classroom proficiency and some definitions.

In the 1980s, the American Council on Teaching of Foreign Languages ACTFL undertook, along with the Educational Testing Service, to develop a global proficiency measure, which was called, not surprisingly, the ACTFL/ETS oral proficiency rating scales. What's interesting about the scales, for those of you who are not foreign languages professionals, is the fact probably, that for the first time, in the history of foreign language teaching in this century, there exists a common metric for the assessment of secondary and post-secondary students, a standardized instrument that allows everyone to agree on what the terms mean. The term "proficient," then, never really meant proficient to do everything all the time, everywhere, in every way possible, but simply to perform certain tasks in certain settings at a certain degree of accuracy as defined by the scales. That doesn't mean that everyone agrees that the scales themselves are perfect. There's no general consensus that this is the only reliable measure. In fact, there's a great deal of debate raging over the content validity of the proficiency scales. But one of the points, which is important for non-foreign language professionals to note, is that this is one instrument that everybody can focus their attention on and begin to talk about as a way of looking at student performance.

I bring that up because in a previous life, one which I enjoyed a great deal and miss a great deal, I worked with ESL and bilingual programs. One of the greatest frustrations was the lack of appropriate instruments to find out what children knew and were able to do. In proficiency testing, one is always focusing on what the learner can do, under what circumstances, and how well. Whereas, when I worked in ESL and bilingual education, I was never quite sure what the tests were really supposed to be testing. One advantage that those who work in the assessment of language minority children should have over foreign language professionals is in the area of identifying goals and objectives. The purpose for assessing English language skills should be to find out if students have acquired the English skills necessary for successful academic performance at or above grade level. In contrast, in foreign language, we very rarely know what students are going to be able to do with their language skills. We don't know the purposes to which they will put their language skills. It's really hard to figure out how to find out what children know when you really don't know what you expect them to know and be able to do in the first place. And if you really don't know what you want them to do, how do you know what to teach them? If you don't know what to teach them, it's awfully hard to decide how to test them. That should not be the case when we work with language minority students, because we are very clear about what we want them to be able to do. We want them to be able to succeed in school. John Oller has said it very well: "Language is the key to successful endeavors, especially, in the school setting." If we know what kids are supposed to be able to do, then why aren't we finding out if they can do it?

It seems to me that entry and exit decisions were based on the wrong things when I worked in ESL/Bilingual education. If you want to know whether a student can perform well academically, the first thing you probably ought to know is: "What are the demands of the academic curriculum from the language perspective?" At every grade level and in every content domain, that may differ; therefore, a student at the third grade level may need to understand this much English, speak this much English, read that much English, write that much English. It might be different for a fourth grader learning social studies or a second grader learning science. Yet, the tests we were working with all looked at students' oral production. Some of them are very discrete point, such as whether a student could discriminate between the sounds of yellow and jello. Except when we teach the concept "matter changes form," we never use the term jello in the third grade. Yet, discriminating yellow from jello was a question on the test, and whether you got to stay in the program depended on whether you understood the difference. That kind of decontextualized assessment of language seemed to be irrelevant to what we needed the kids to be able to do.

As an ESL program director, I was always terrified when the children took their language proficiency tests. Part of me really wanted them to do well, because that was what our program was all about—helping them to succeed. We wanted every child to do as well as possible. But this little voice inside of me said, "Oh, if they do well, we won't be able to help them anymore." Because no matter what the tests said, I knew that some of those children weren't quite ready to make it on their own. It seemed to me an awfully silly way to make a decision about who gets in, who stays in, and who gets out.

It's all that which brings me to my central argument.

The most promising way, then, to address the concern of the appropriate assessment language proficiency is through instructionally-based assessments, such as the ones we have been hearing about at this symposium and certainly the ones I think are relevant from my experience with foreign language immersion programs. In foreign language immersion programs, students learn content through a language in which they have limited skills. Immersion teachers are responsible for ensuring that their students achieve the objectives of the school curriculum while gaining skills in a new language. In this respect, their roles and responsibilities parallel those of teachers who work with language minority students.

For the last four and a half years, I have been involved in a project to identify the training needs of foreign language immersion teachers and to help develop training materials to meet those needs. I would be the first to say, and I really want to stress this, that foreign language immersion is not the same as ESL or bilingual education, (nor should it be), but the needs of the teachers who work in these fields are similar in that they're all engaged in teaching content in a language that is new to their students. Also, some of you who are working in the field of developmental bilingual education may find it interesting to hear some of the training issues that are involved in foreign language immersion. We have been helping teachers learn how to teach in these foreign language settings and to find out, indeed, if children are learning. In this project, I have come to believe that the teaching of language and content should be inseparable. I am going to say that again, because I think that is the most important thing I have to say today. The teaching of language and content ought be inseparable. Language is learned best through a context and a content, particularly when the aim of the language program is to enable students to be successful academically in their new language. John Oller has just told us that language is important to all educational endeavors, and that to separate language from meaning, language from thought and cognition, and from content, is to make a mockery of the business that we're all about.

Language objectives and content objectives must be tied to one another. Both sets of objectives must be considered when planning for teaching and when planning for testing. We tell our teachers that planning for testing takes place at the time that you plan for your teaching. Teachers must identify the language demands of the curriculum and plan to include means for students to gain in language as they grow in concept attainment. Anne Snow, Fred Genese, and I have suggested elsewhere a model for the integration of language objectives with the teaching of content, and visa versa, and have demonstrated how the roles of the ESL, the bilingual teacher, the mainstream teacher, and the foreign language teacher are fulfilled within that framework. I'm not going to go into that paper here, but I do want to stress the importance of teaching language through content and the importance of considering every content lesson a language lesson as well.

Teaching and testing go hand and hand. As John points out in his paper, (a point he didn't mention this morning), testing activities should be as broad as our teaching activities. In fact, planning for testing and planning for teaching need to be done at the same time. Effective foreign language immersion teachers begin to plan by first thinking about what they want students to learn. Then, when they know what they want students to learn, they've got to figure out how they're going to know it when they see it. If they know what they want children to learn and how they're going to find out if they have learned, then the next step also falls in line, which is, how you're going to get children ready to show you or to perform their knowledge. Those are the enabling activities; that's the teaching part. So, learning and teaching and testing all belong together. Good immersion teachers, then, are able to ensure that their objectives, their teaching, and their testing all fit together, because they see them as inextricably tied to one another. And, they define their objectives, teaching, and testing both in terms of content and language.

Since teaching concepts in a new language often requires that immersion teachers use visual and other concrete experiences during instruction, it follows that similar approaches are appropriate when testing students. Students should have access to materials that help them show the teacher what they know, even when they can't always tell her.

In assessing students, immersion teachers are most concerned with finding out what students have learned and allowing students to demonstrate what they have learned. The emphasis is on what students can do and do know, not on what they don't know and can't do. John told us in his paper that what we need most in our profession are integrative tests that tie teaching to learning to assessment. In foreign language immersion, classroom-based language assessments that are conducted as part of the instructional delivery system serve a number of important masters. First, they seem to be the most appropriate way of finding out whether students have the language skills needed for academic performance, precisely, because the assessment ties language to its purpose, which is content learning. These assessments are authentic in that they measure student proficiency in the real contexts in which language use occurs. They're integrated and assess the range of skills needed in the classroom for successful academic performance. These tests, in essence, have content validity, because, (as I heard the term used yesterday), they "test the right thing." Classroom-based performance assessments put the focus where it belongs, on student growth. Performance assessments such as portfolios, systematic observation, and teacher evaluations of student products and projects are effective ways to find out about student progress in relation to the objectives we've set for them. Because they're based on student performance; they show us what students can do and do know, and they compare each student to his or her last performance, they only compare students to themselves, not to some idealized and probably non-existent average student or native speaker.

Last, they're the most appropriate way of ensuring that the delivery of instruction is commensurate with the linguistic proficiency of the student at that point in time and in that content domain.

From the day to day instructional perspective, the marriage of language assessment with content assessment helps teachers, whether they're foreign language immersion teachers or those who teach language minority students, engage in a constant formative diagnostic feedback loop. In our training of foreign language immersion teachers, we emphasize the importance of surveying students' background knowledge prior to introducing a new concept. Every teacher does that but, for foreign language immersion teachers, this also means that they must know the range of the students' linguistic ability to handle the concepts. The teacher needs to know the language demands of the curriculum objectives and the extent to which special strategies, manipulatives, and concrete materials will be necessary for instructional delivery.

Immersion teachers are content teachers, but they're also language teachers. We believe that every content lesson should be a language lesson as well, and that foreign language immersion teachers need to plan as conscientiously for language growth as they do for content. In part, planning for language growth means the teacher must be continuously assessing where students are in relationship to where they ought to be and using that assessment data to identify areas where further development of language growth is needed.

It's clear, then, that as instruction progresses and as teachers observe the growth of students, a great deal of assessment data can be collected about the achievement of both content and language objectives. These data provide important information about each individual student but, in the aggregate, data from systematic observations, checklists, portfolios, and teacher-made tests also provide information about the effectiveness of the instructional program.

In conclusion, trends in foreign language teaching and testing have two major implications for the assessment of language minority children. One is, perhaps, a different definition of proficient -- a recognization that all language users, both native and non-native, are differentially proficient to perform language tasks in different settings and at varying levels of performance. None of us is completely proficient, and this definition of proficiency renders the notion of limited proficiency almost meaningless, as a system of categorizing learners. Language minority students bring with them a rich resource in their home language and culture. The label, "limited proficiency students," as John tells us in his paper, only perpetuates a deficit model of instruction and relegates ESL and bilingual education to a compensatory role. Perhaps a more useful way at looking at proficiency, as in the newer definition in foreign language, is to describe what learners can do, under what circumstances, and how well. For language minority students, defining proficiency in terms of classroom language -- the tasks, the functions, the contexts, and the contents in which they must perform -- will allow us to focus assessment measures where they belong, on academic performance. The second implication of foreign language immersion is that the teaching and testing of English in ESL bilingual programs must be integrated with the content students are to learn. If the teaching and testing of English were more intimately tied to the learning of content, we might more effectively integrate teaching, learning, and assessment. To paraphrase the late Ron Edmonds, (and I'm sure many of you have heard this before) "All children can learn. All children must learn, and all teachers must learn to teach (...and I'll throw in my paraphrase, and equitably assess) all children."

Performance Assessment of Language Minority Students

Jack S. Damico
University of Southwestern Louisiana

Introduction

Performance assessment of language minority students is a complex process that requires the application of theoretically defensible procedures that are carefully designed and systematically implemented. Due to the differences between language minority students in the schools and those ESL/EFL students typically studied by language testing researchers, performance assessment in the schools must involve the utilization of procedures that are more authentic, more functional, more descriptive, and more individualized than those typically recommended by second language testing researchers. This paper proposes a descriptive approach to performance assessment that is theoretically defensible and psychometrically sufficient. The characteristics necessary for successful performance assessment, the assessment process, and actual assessment techniques are discussed.

> "To me it seems to be generally desirable in instructional contexts to focus tests diagnostically only against a contextual backdrop where attention is directed toward comprehending or producing meaningful sequences of elements in the language" (Oller, 1983:354).

When I first read this passage in a working manuscript of John Oller's 1983 chapter, "A consensus for the eighties?", it had a galvanizing effect on me. This suggestion of a "pragmatic" approach to assessment was the final impetus for me to shift my theoretical stance and my practices involving performance assessment of language minority students in the schools. I recognized that although I was very interested in the excellent work going on in second language testing research, my research involving language minority students in the schools required something different.

While there are a number of purposes for evaluation and assessment of language minority students in the schools (Henning, 1987), those purposes most relevant to my concerns revolved around the student and the student's ability. Using Oller's conceptual writings as guidance, my work has focused on ways to provide a rich description of the individual student's communicative performance and that student's underlying language proficiency. To accomplish these objectives, the assessment procedures designed and implemented for these students have to be more **authentic**, more **functional**, more **descriptive**, and more **individualized** than those procedures recommended in the second language testing literature. This is due to the differences between the language minority students my research targeted and the ESL/EFL students discussed by many other language testing researchers.

The Differences Between The Populations

The second language testing literature typically focuses on students who are enrolled in ESL or EFL classes. These students are tested to determine placement in second language classrooms or to determine their progress in these classes. With much of this research, the students are usually older than elementary level and/or there is an assumption that these students have normal language-learning abilities. The situation facing the typical language minority student in our public schools -- particularly at the elementary and mid-school level -- is very different.

Unlike the majority of students discussed in the second language testing research literature, the language minority students usually targeted for evaluation in the public schools are located in environments that are not conducive to language diversity. They are frequently the only students in their classrooms who are non-English speakers. Additionally, their teachers are unlikely to have knowledge of their first language or even of the process of second language acquisition. As a result, normal acquisitional phenomena may be viewed as an indication of language-learning problems (Hamayan & Damico, 1991). Within this environment there may even be prejudice toward students who are speakers of other languages (Ogbu, 1978).

Second, unlike older ESL/EFL students, the language minority students are typically compared in their routine academic and conversational performance with the mainstream students rather than other language minority students. In such cases, they naturally perform more poorly since they don't have the same proficiency in English and, because they are not being tested and compared in the ESL or EFL classroom with their peers, their performances are always suspect when they perform below the mainstream expectations (Cummins, 1984; Ortiz & Wilkinson, 1987).

Third, many of these language minority students come from what Ogbu (1978) has termed the "caste minority" group. That is, a group of individuals that may or may not have been born in this country but who are usually regarded by the mainstream and dominant population as being inferior. As a result, these individuals are **perceived** as being less-intelligent, less-motivated, and less-able to match the mainstream students in a range of activities. These biased perceptions give rise to lowered expectations and frequently result in a disempowerment of these students that is manifested not only in their academic performances but in the ways that they perform and are evaluated during assessment (Cummins, 1986; 1989; Mercer, 1984; Ogbu, 1978).

The fourth difference between these students only heightens the perceptions created by the first three differences: these students are usually individuals that do not have well-documented first language proficiency. Rather than having demonstrated their first language through performances that enabled them to enter into a second language learning context (e.g., an EFL class), many language minority students are kindergartners or first graders who are still acquiring their first language. Consequently, little is known about their native language proficiency. Even when these students are older, they are frequently recent immigrants that have few academic records from their home countries that could document their performances (Cloud, 1991). As a result, there is no assumption of sufficient first language proficiency or normal language-learning capacity. Rather, given the first three conditions, these students may be suspected of poor language proficiency in both their first and second languages. That is, they may be suspected of exhibiting a language-learning impairment.

The final difference is a natural consequence of this suspicion and it makes the situation for many language minority students most desperate: the purposes of assessment are frequently different. Unlike the ESL/EFL students who are assessed for placement in classes to supplement their normal language-learning proficiency with the addition of a second language, language minority students may be assessed to determine remedial placements or placement within special education programs. A harsh reality in our public schools is that many language minority students are mis-diagnosed and enrolled in special education programs or remedial tracks that reduce their academic potential as normal language-learners (Cummins, 1984; Fradd, 1987; Oakes, 1985; Ortiz & Yates, 1983).

As a result of these differences, performance assessment for many language minority students requires a different focus. Not only must the usual testing purposes be accomplished (O'Malley, 1989), the evaluator must also be able to address specific questions regarding an individual's underlying language proficiency and learning potential. In the remainder of this paper, a descriptive approach with a pragmatic focus that has been effective in the performance assessment of language minority students will be detailed. Although this descriptive approach is aimed at actual evaluation and diagnosis of individual students for selection and placement purposes, it can also be utilized in program evaluation for formative and summative evaluation purposes (Navarrete, Wilde, Nelson, Martinez, & Hargett, 1990).

Descriptive Performance Assessment

To adequately evaluate language minority students in the schools, performance assessment practices must be consistent with the currently accepted theoretical construct of language proficiency (Bachman, 1990a; Oller, 1989; Oller & Damico, 1991) and they must be carefully designed to meet the numerous assessment requirements within the public school environment. The theoretical grounding will help ensure strong validity indices while the design characteristics will aid in the construction of assessment procedures that possess sufficient reliability and educational utility.

The Construct of Language Proficiency

Since 1961 when John B. Carroll raised concerns regarding the artificialness of the "discrete-point" testing methodology, numerous second language researchers have advocated assessment and evaluation procedures that stress the interrelatedness of language as a psychological construct. Calling for the development of "integrative" (Briere, 1973; Oller, 1972; Spolsky, 1973; Upshur, 1973), "pragmatic" (Oller, 1979; 1983), "edumetric" (Cziko, 1983; Hudson & Lynch, 1984), "communicative" (Bachman, 1990a; Canale, 1987; 1988; Olstain & Blum-Kulka, 1985; Shohamy, 1991; Van Lier, 1989; Wesche, 1987) and "informal" (Brindley, 1986; Navarrete, et al., 1990) assessment procedures, these researchers utilized the most defensible constructs of language proficiency available to them to justify their test design. Descriptive performance assessment is also based on the most defensible construct of language proficiency available.

Currently, language proficiency is viewed as a componentially complex psychological construct with a powerful synergistic quality that enables language or communicative ability to act as a coherent and integrated totality when it is manifested in performance (Bachman, 1990b; Carroll, 1983; Oller, 1983; 1989). While there are several models or frameworks that are consistent with this construct (e.g., Bachman, 1990a; Canale & Swain, 1980; Cummins, 1984; Shohamy, 1988), the hierarchical model of language proficiency proposed by Oller (1983; 1989; Oller & Damico, 1991), seems most appropriate and is utilized for the design of this descriptive approach to performance assessment.

In this hierarchical model, language proficiency is recognized as a multicomponential and generative semiotic system that functions in an integrated fashion in most communicative contexts. For practical purposes, language proficiency exist when the individual components (e.g., syntax, morphology, phonology, lexicon) function as an integrated whole. Further, in keeping with the synergistic perspective, this integrated whole is unpredicted by the behavior of the individual components when they are described separately (Fuller, 1982). This is because these separable components are more apparent than real; they are essentially terminological distinctions created in the mind of the linguist or evaluator for ease of discussion and analysis. As noted previously, language is a generative system that exists for the transmission and coding of meaning and these components are aspects of this process. However, they are not divisible and discrete in their functioning; they function holistically. Consequently, when observable aspects of these components are isolated in artificial tasks during assessment, the tasks are not assessing language or communication but some splinter skill quite different from true language proficiency.

Another facet of this synergistic quality of the hierarchical model is that language proficiency is not viewed as an autonomous semiotic system. It is an integrated system that is intimately influenced by other semiotic and cognitive systems and by extraneous variables. Consequently, performance is highly influenced by factors like memory, perception, culture, motivation, fatigue, experience, anxiety, and learning. For effective and valid assessment, therefore, it is important that language and communicative behavior be assessed in natural contexts.

As discussed elsewhere (Damico, 1991a; Oller & Damico, 1991), reliance on Oller's hierarchical model results in a number of advantages when trying to account for data and concepts reported in the second language literature. Such discussion, however, is beyond the scope of this paper. For our current purposes, this model helps in the design of an effective descriptive performance assessment system.

The Design Characteristics of Descriptive Assessment

Based on the hierarchical model regarding the construct of language proficiency and on the specific purposes for performance assessment previously discussed, there are several design characteristics necessary to the development of effective procedures for descriptive performance assessment of language minority students (Damico, 1991a; Damico, Secord & Wiig, 1991). These characteristics concern the authenticity of the data collected and analyzed, the functionality of the behaviors evaluated, a sufficiently rich description of language proficiency to accomplish assessment objectives, the necessity for a focus on each individual being assessed, and the assurance of psychometric veracity (See Table 1).

Table 1
Design Characteristics Necessary for Descriptive Assessment

- Authenticity of the Collected Data
  - Linguistic Realism
  - Ecological Validity

- Focus on Functionality
  - Effectiveness of Meaning Transmission
  - Fluency of Meaning Transmission
  - Appropriateness of Meaning Transmission

- Rich Description of Language Proficiency
  - Descriptive Analysis
  - Explanatory Analysis

- Emphasis on the Individual

- Assurance of Psychometric Veracity
  - Reliability
  - Validity
  - Educational and Programmatic Utility

Authenticity

The first characteristic necessary to effective descriptive assessment relates directly to the synergistic quality of language proficiency. Since language proficiency is synergistic in terms of its internalized structure, its relationship with other semiotic and cognitive abilities, and its interaction with external variables, it is not possible to assess language and communication apart from the influence of intrinsic cognitive factors and extrinsic contextual features. Consequently, assessment must be structured to observe language during actual communicative activities within real contexts. The language and communication behaviors assessed must be authentic (Damico, 1991a; Oller, 1979; Seliger, 1982; Shohamy & Reves, 1982).

For our purposes, authenticity means that the methods used in assessment focus on data that possess linguistic realism and ecological validity (see Table 1). Linguistic realism requires that assessment procedures treat linguistic behavior as a complex and synergistic phenomenon that exists primarily for the transmission and interpretation of meaning (Crystal, 1987; Oller, 1979; 1983; 1989; Shuy, 1981). The task of assessment, therefore, is to collect data that are meaning-based and integrative rather than data that attempt to fragment language or communication into discrete points or components. Consequently, the data of interest in assessment should be actual utterances and other meaningful "chunks" of linguistic behavior that serve to transmit an idea or intention from a speaker or writer to a listener or reader (Damico, Secord, & Wiig, 1991). In this regard, there should be a focus on the analysis of discourse.

Ecological validity is another aspect of authenticity that must be accommodated during assessment. Numerous researchers have demonstrated that the behaviors manifested during isolated and artificial testing procedures are unlike the linguistic and communicative behaviors noted in real language usage situations (Carroll, 1961; Cummins, 1984; Douglas & Selinker, 1985; Shohamy, 1983). Rather than trying to isolate the assessment process from contextual influence, therefore, assessment should be accomplished in naturalistic settings where true communicative performance is occurring and is influenced by contextual factors. Such practices will enable the evaluator to discern the effects of contextual variables on the student's communicative performance and will enable assessment to remain consistent with the emphasis on relativism in behavior analysis (Kagan, 1967; Oller, 1979).

Oller's (1979) work regarding the development of pragmatic assessment procedures is consistent with this focus on authenticity. By incorporating the work of several other language testing researchers, Oller suggested that language testing adhere to three "pragmatic criteria." The first, that data be meaning-based, required that data be collected from tasks that were motivated by a desire to transmit meaning or achieve comprehensibility during meaning transmission. The second, that data be contextually-embedded, required that the data under scrutiny be produced in a contextually rich environment. The third criterion, that data be temporally-constrained, required that tasks used to collect the data fit into the normal temporal envelope of communicative interaction. Taken together, these three pragmatic constraints act to ensure linguistic realism and ecological validity.

Functionality

The second characteristic necessary for effective descriptive assessment is concerned with how performance is evaluated or, put in more operational terms, what should be measured. In the second language testing literature, there have been a number of attempts to identify the various components of language proficiency and then design tools or procedures to measure these components (Bachman, 1990a; Bachman & Palmer, 1982; Canale, 1983; 1987; Harley, Cummins, Swain, & Allen, 1990; Swain, 1985). This research has been controversial and no clear consensus has emerged regarding what components to measure and how to measure them. As discussed by Oller (1979; 1989; Oller & Damico, 1991), this is probably due to the powerful integrative trait manifested by language proficiency during performance. Division into components should not occur during data collection or preliminary analysis. Such strategies strip the synergy inherent in language behavior. Rather, there should be an initial focus on language proficiency and communicative performance from a functional perspective. (Attention to more traditional linguistic componential perspectives should be reserved for the stage referred to later as "explanatory analysis").

The focus on functionality suggests that instead of testing a student's knowledge of discrete points of superficial language structure -- or even the student's ability to effectively demonstrate separate componential knowledge of strategic or grammatical competence -- in order to indicate potential language or communicative difficulty, the evaluator asks the question, "How successful is this student as a communicator"? This question of success is based on how well the student functions on three criteria: the effectiveness of meaning transmission, the fluency of meaning transmission, and the appropriateness of meaning transmission (see Table 1).

The Effectiveness of Meaning Transmission. This criterion relates to the primary goal of communication: the formulation, comprehension, and transmission of meaning. Since language is a semiotic system that exists to achieve an understanding of what occurs in the world and since some aspect of this understanding is formulated into communication to relate that understanding to others, how well the individual handles this message (either as a speaker or hearer) is directly relevant to that person's success. The key element, of course, is the message and achieving comprehensibility so that the message is transmitted. Using a functional focus, if the meaning is transmitted -- regardless of how that transmission is achieved -- then communication is accomplished and the individual is effective.

The Fluency of Meaning Transmission. From a functional performance perspective, however, success is more than just getting the meaning across. As stressed by Carroll (1961), the fluency of the interaction must also be considered since successful communicators must be able to formulate, transmit, or comprehend the message within the temporal constraints of communicative interaction. If a student's communicative attempt is delayed, then the flow of communication is affected and this will result in a devaluation of the individual's rating as a communicator. Additionally, a successful communicator can also repair an initial interaction if meaning transmission is not successful. As a speaker, can the student reformulate the message so that it is better comprehended by others? As a listener, can the student successfully ask for clarification or effectively utilize contextual cues if the initial message is incomprehensible? This ability to seek clarification is another facet of the fluency of meaning transmission.

The Appropriateness of Meaning Transmission. The third criterion for success as a functional communicator is to accomplish the first two objectives in a manner appropriate to the contextual constraints in which the student is immersed at the time of the interaction. Realistically, language and communication are significantly influenced by the expectations that members of a linguistic community share regarding their communicative norms. The attitudes that individuals form and the opportunities afforded to individuals in that community are frequently dependent on those expectations. When addressing the needs of language minority students, this criterion is very important. On numerous occasions, language minority students are poorly evaluated not because of an inability to transmit meaning, but to do so in a culturally appropriate manner (Cummins, 1984; Hamayan & Damico, 1991; Iglesias, 1985).

These three criteria enable the descriptive assessment process to focus on the functional dimension of communicative ability in a manner that transcends the need to divide language proficiency into a variety of skills, modules or components at the time of assessment. Consequently, the synergistic quality of language so important within communicative settings is preserved. Additionally, this initial focus on functionality enables the evaluator to answer real and pragmatic questions about the minority language student's ability to function in the second (and the first) language context as a successful communicator. Actual techniques that can be utilized to accomplish this functional focus will be discussed below.

Descriptiveness

The third essential characteristic of descriptive assessment involves the purposes of evaluation and the types of analyses performed to achieve these purposes. As previously discussed, many language minority students are not only assessed for selection and placement in bilingual programs or for limited proficiency instruction in English. Many of these students are also assessed to determine their underlying language-learning proficiency for placement in special education or remedial programs. It is important, therefore, that descriptive assessment focus on two objectives. First, in order to meet the needs of regular bilingual education programs, governmental funding requirements, and legal regulations (O'Malley, 1989), assessment should provide a detailed description of the individual's communicative performance in English. Second, this descriptive information must then be used to comment on the student's underlying language proficiency. That is, the first objective of descriptive assessment is to determine how successful the student is as a communicator in English and then, if the student is not successful, to determine the reasons for this lack of success. To accomplish these objectives, the descriptive process must utilize a **bi-level analysis paradigm** that incorporates an initial descriptive analysis of communicative performance with a detailed explanatory analysis of language proficiency (Damico, 1991a) (see Table 1).

Descriptive Analysis. At this descriptive level of analysis, the evaluator typically uses actual descriptive assessment procedures to observe the student's communicative English performance in the contexts and modalities of interest and to determine whether or not the student is communicatively successful. This determination is made by asking the question presented in the previous section on functionality, "How successful a communicator is the student in the context and modality of interest"? This question is answered by one of two strategies. In the first strategy, the descriptive procedure is designed to focus directly on observable behaviors that have been found to be necessary for successful communication in the targeted context and modality. These behaviors are usually selected while designing the descriptive assessment tool through the application of criterion-based, communication-based, or curriculum-based procedures (Cziko, 1983; Hudson & Lynch, 1984; Marston & Magnusson, 1987; Nelson, 1989; Tucker, 1985).

The second strategy used to answer the functionality question at the descriptive level of analysis focuses on potential problematic behaviors. In this strategy, descriptive procedures are designed to detect behaviors that are believed to be valid indices of communicative difficulty. Specific behaviors are identified as indicating when a student is experiencing problems during the communicative interaction and these behaviors are targeted for coding during assessment (Damico, Oller & Storey, 1983; Goodman & Goodman, 1977; Mattes, 1985).

Regardless of the strategy used, this descriptive level of analysis provides a determination of the student's success as a communicator in the targeted English contexts. It is at this level that the primary assessment objectives discussed by Henning (1987) and O'Malley (1989) for limited English proficient (LEP) student identification, placement, and program evaluation are accomplished. By using the actual descriptive analysis procedures, data are provided to document the student's strengths and weaknesses in English, the student's overall success as an English language user, and even the student's progress over time (when pre/post analyses are conducted). Many of the descriptive analysis procedures provide objective scores to rank individual students for placement or reclassification purposes or, if such scores are not available, scoring adaptations like those suggested by Navarrete et al. (1990) may be applied.

It should be noted that during this descriptive level of analysis, communicative performance should be analyzed in different contexts and in different modalities to ensure a rich and wide-ranging coverage of the minority student's success in various language usage manifestations. As discussed by Cummins (1984), there are various dimensions of language proficiency that interact to produce performance distinctions in bilingual (and monolingual) language users (e.g., the CALP and BICS distinction). Similar recommendations for the assessment of language performance in different modalities and contexts have been advocated by others (Canale, 1983; Damico, 1991a; Luria, 1981; Oller, 1979). A discussion of several descriptive analysis procedures is provided below.

If there are no difficulties noted after a descriptive analysis is performed or if there are no attempts to refer the language minority student for further (special education or remedial) testing, then the student is considered for appropriate placement in a bilingual program, a program for English instruction, or in the mainstream classroom. These placements are dependent on how successful the student is as an English language user in the contexts and modalities of interest. This decision meets the general student assessment purposes discussed by Henning (1987) and O'Malley (1989). If there are difficulties indicated at this level of analysis and if there are concerns regarding the student's language-learning proficiency, then the second level of analysis is necessary.

Explanatory Analysis. This analysis seeks to determine the causal factors for the communicative difficulties noted in the descriptive analysis. At this analytic level, the examiner notes the absence of the indices for success (the first strategy above) or the presence of the problematic behaviors (the second strategy above) and seeks to determine why these behaviors occurred.

The explanatory level of analysis typically does not involve additional data collection or assessment procedures. Rather, this level involves a deeper interpretation of the data collected at the descriptive level. The evaluator attempts to explain what aspects of the context, the student's social/cultural experience, or the individual's cognitive abilities or linguistic proficiency can account for the described difficulties. At this level, the evaluator attempts to determine the adequacy of the student's underlying language proficiency or comment on the effectiveness of the student's deeper semiotic capacities (see Oller & Damico, 1991). To accomplish this analysis, a number of strategies can be utilized (Damico, 1991a; Goodman & Goodman, 1977). One effective way to structure this analysis for language minority students is with a set of questions that may be systematically applied to explanatory analysis. A modification of Damico's list (1991a) will be discussed below.

The descriptiveness characteristic enables this performance assessment approach to accomplish the various purposes of assessment while maintaining a functional perspective with authentic language and communicative data. In order to fully adhere to each of these three design characteristics, however, the fourth characteristic of descriptive performance assessment is required. That is, the assessment must be individualized.

Individualized Assessment

Descriptive performance assessment requires individualized assessment. Unlike a number of the discrete point language tests and even some of the more integrative testing procedures, the descriptive process emphasizes individualized observation and analysis. This characteristic is particularly important when determining the need for an explanatory analysis. While the description of the student's communicative performance from an authentic and functional perspective is difficult enough, the analysis of the student's underlying language proficiency based on results of the descriptive analysis procedures is virtually impossible unless conducted on one student at a time. In order to richly describe the complex phenomenon of language proficiency for the purposes previously discussed, time and effort are required. Given the importance of the placement decisions, however, such individualization should not seem excessive. Quality language education of language minority students should require nothing less.

Psychometric Veracity

The final essential characteristic of descriptive performance assessment is psychometric veracity. Similar to a general definition of construct validity (Cronbach, 1971; Messick, 1980), this characteristic reflects the interaction of the other four characteristics once they are carefully implemented. This concept embraces the idea that the tests and procedures used during assessment must be genuine and effective measures of language proficiency and communicative performance. Consequently, veracity requires strong psychometric qualities of reliability, validity, and educational or programmatic utility. In order to exhibit veracity, the assessment procedures must focus on authentic data and must target specific behaviors to use as indices of language proficiency and communicative performance. The evaluator must know what behaviors indicate successful or unsuccessful communication and these behaviors must be able to reflect on the student's underlying language proficiency. The identification of the indices during test development involves several steps (see Table 2).

Table 2
The Steps Required for Determination of Valid Behavioral Indices

Step 1. Select the targeted contexts and modalities.

Step 2. Identify functional behaviors required for meaning transmission.
- Strategy One: Identify behaviors that are indices of successful communicative performance.
- Strategy Two: Identify behaviors that are indices of communicative difficulty.

Step 3. Determine the Reliability of the selected behaviors.
- Temporal reliability
- Interexaminer reliability

Step 4. Determine the Validity of the selected behaviors.

Step 5. Determine the Educational and Programmatic Utility of the selected behaviors and the assessment procedure.

First, the test designer or the evaluator must determine the actual contexts and communicative modalities that will be targeted by the descriptive procedure. This is essential since the data targeted and the student's manifestations of language and communicative ability will differ across contexts and modalities. For example, the language needed to be successful during a writing lesson in the classroom is different from the language needed for a conversation at a friend's home after school. The proficiency required and the communicative behaviors manifested are quite different (Cummins, 1984). To identify valid indices, therefore, the context and modality of interest must be selected.

Once the target manifestations are selected, then behaviors that have a functional role in the transmission of meaning in that manifestation should be identified. This identification is the second step and the two strategies that may be used to accomplish it have been previously discussed. The third step involves the psychometric concept of reliability. Reliability is necessary to ensure that the specific behaviors used as indices for language proficiency are consistent and stable enough in their occurrence (temporally reliable) to be considered as true indices of a stable underlying language proficiency. Additionally, these behaviors must also be easy to observe and code by individuals trained to use the procedures. If different evaluators cannot easily code these behaviors and agree on their occurrences (interexaminer reliability), then the potential of these behaviors as effective indices is greatly diminished. If these behaviors are demonstrably reliable over time and across examiners, however, then the next step can be taken with these behaviors.

A determination of the validity of these behaviors as indices is the next step. Simply put, how well do these reliable behaviors actually reflect on the language minority student's underlying language proficiency? If a descriptive procedure using these behaviors enables an examiner to make accurate predictions about an individual's language-based performance over time and outside of the assessment situation, then validity is demonstrated and the descriptive procedure is useful (Bachman, 1990a; Cronbach, 1971; Oller, 1979). Without some indication of a procedure's validity and an index's role in that validity, however, the effectiveness of the procedure is reduced.

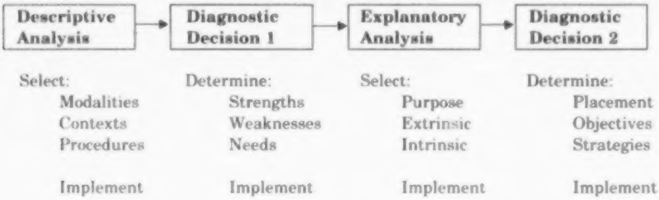
The last step in establishing psychometric veracity involves determining the educational or programmatic utility of the procedures. Regardless of how reliable and valid a procedure (and its indices) might be, it must be relatively easy to learn and apply in the educational setting or instructional program and it must reflect the instructional goals and objectives of the setting or program. If a procedure requires an inordinate amount of time or equipment to implement or if the data obtained is inconsequential to the setting or program, then it is unlikely that the procedure will be embraced by any evaluator constrained for time. Descriptive performance assessment procedures must reflect the realities and limitations of the school systems that employ the evaluators.

These five characteristics of descriptive performance assessment act to ensure that the processes and procedures used to evaluate language minority students are effective. To implement performance assessment, however, it is not enough to merely describe the characteristics of the descriptive approach. A description of the actual process of assessment and some of the procedures useful for performance assessment is also necessary.

The Descriptive Assessment Process

As previously discussed, an evaluator using the descriptive approach conducts the assessment process differently from other assessment approaches. Remaining consistent with the authenticity characteristic, communication is assessed as it functions holistically in its various manifestations and within naturalistic contexts. The process for a complete assessment involves four sequential stages. First, the evaluator conducts a descriptive analysis. Based on the findings of this analysis, the second stage involves making the first set of diagnostic decisions. If assessment is still warranted after this stage, the next stage of assessment involves the utilization of an explanatory analysis. Finally, based on the results of this deeper level of analysis, the second set of diagnostic decisions is made. Figure 1 provides a flowchart description of this descriptive assessment process.

Figure 1
A flowchart description of the
Descriptive Performance Assessment Process



When conducting a **descriptive analysis**, the evaluator asks the question, "In the present domain of interest, how successful is this student as a communicator?" To conduct this stage of the process, the evaluator determines the modalities of language use and the observational contexts that should be assessed. The evaluator then chooses communicative assessment procedures that will allow the description of communication as its functions in whatever modalities and contexts are appropriate to the objectives of the program or setting. These tend to vary according to the individual programs. For example, Damico (1991a) has specified three manifestations of language use that are important to his assessment objectives. In other situations, however, only one manifestation of language use might be important. Regardless of whether several modalities (e.g., writing, speaking, reading) or several contexts (e.g., conversation with friends, job interview, classroom discussion group, lesson recitation) are evaluated, assessment procedures that describe authentic communication from a functional perspective are required for evaluation. These assessment procedures should typically focus on all aspects of communicative effectiveness (language and speech) together and allow for a determination of communicative success based on the three criteria of effectiveness, fluency, and appropriateness of meaning transmission.

Once the stage of descriptive analysis is completed, the evaluator makes the **first** set of **diagnostic decisions**. At this time, the evaluator uses the data collected during the previous stage to determine whether or not there are any communicative difficulties in the targeted language manifestations. If no difficulties are noted, then the assessment process is completed and the evaluator can describe the individual's strengths as evidence of communicative success and strong language proficiency. If difficulties are noted in one or more of the language manifestations, however, then the evaluator has several tasks.

First, from a functional perspective, the evaluator describes the student's individual strengths and weaknesses. This information will assist in identification and placement into supportive educational programs. Next, the evaluator determines the actual needs of the student at this stage of the assessment process. If there is evidence of strong first language ability or if there is no desire to refer the student on for further (special education) assessment, then the actual placement decisions (e.g., bilingual classroom, ESL instruction) can be made for this student and the evaluation is completed. However, if there is no evidence of normal first language ability, if there are indications of potential language-learning difficulties at the deeper level of language proficiency or semiotic capacity, or if others in the educational settings request further evaluation, then the next stage of the descriptive assessment process occurs.

This stage of the descriptive process involves the second level of analysis described under the bi-level analysis paradigm mentioned previously -- **explanatory analysis**. At this stage of assessment, the evaluator examines the difficulties noted during the descriptive analysis stage and attempts to determine why the individual had these particular difficulties in the manifestations under scrutiny. Initially, extraneous variables are examined as potential explanations for these problematic behaviors (e.g., second language acquisition phenomena, contextual complexity, listener reactions, significant cultural differences). If no extraneous explanatory factors are noted, then a more systematic analysis of the actual linguistic data is initiated. Based on this analysis, the evaluator determines the underlying causes of the problematic behaviors identified during the descriptive analysis stage and can determine if a true intrinsic language-learning disorder exists.

Finally, at the last stage of the process, the **second** set of **diagnostic decisions** are made. Based on the results of the explanatory analysis and the opportunities available to the student, appropriate placement recommendations are made. If the explanatory analysis demonstrates difficulties due to extraneous variables and not due to the student's underlying language proficiency or deeper semiotic capacity, then the recommendations will be for various types of support systems or programs that will benefit the student's acquisition of English and academic material within regular educational formats. A number of authors have discussed various pedagogical strategies along these lines (Chamot & O'Malley, 1987; Cochran, 1989; Crandall, Spanos, Christian, Simich-Dudgeon, & Willetts, 1988; Fradd, 1987; Hamayan & Perlman, 1990; O'Malley & Chamot, 1990). If the difficulties appear to be due to more intrinsic cognitive, linguistic, or semiotic factors, then appropriate placement and remediation in special education may be warranted (Cummins, 1984; Damico & Nye, 1990; Willig & Ortiz, 1991).

In summary, the assessment process utilizes authentic and functional procedures to analyze an individual's communication in a descriptive and pragmatic fashion through the application of a four stage process centering around a bi-level analysis paradigm. This process is consistent with the hierarchical and synergistic model of language proficiency proposed by Oller and may be used to fulfill the purposes of assessment for language minority students.

The Descriptive Analysis Procedures

To remain consistent with the descriptive assessment approach, the assessment tools and procedures utilized to describe language and communication (i.e., procedures implemented during the descriptive analysis stage), must be able to analyze language use in authentic situations from a functional perspective. Over the past 15 years, a number of procedures and tools have been developed that fit these requirements. In general, these procedures can be conveniently organized according to their different data collection formats and the primary behavioral manifestations that these procedures target. In keeping with Cummins (1984), Canale (1984), and Damico (1991a), these tools and procedures will be organized into four major data collection formats (probes, behavioral sampling, rating scales and protocols, and direct and on-line observation) and two general language usage manifestations (conversational and academic). While these divisions are too general to provide a rigorous classification system, they will permit sufficient organization for this discussion.

Probe Procedures

The probe format is the most widely used of the four data collection strategies. It has been used in the design of norm-referenced tests and many integrative and descriptive procedures. There are numerous variations within this category. For example, there are picture elicitation probes, question and answer probes, elicited imitation probes, interactive computer probes, direct translation probes, and role-playing activities to name a few. Probes are structured tasks or activities that elicit a specific language behavior from the individual being assessed. With probes, the evaluator may anticipate the type of response that will be elicited from the student. This is because the task performed, whether discrete-point, integrative, or pragmatic, is carefully designed to elicit a specific behavior.

For conversational purposes, there are a number of probe activities that have been suggested. Brinton and Fujiki (1991), for example, have structured an assessment activity used to probe a

168

student's ability to revise utterances, maintain topics, and ask relevant questions during conversation. Olshtain and Blum-Kulka (1985) have suggested several elicitation techniques that focus on language usage, while Deyes (1984) and Jonz (1987) have adapted cloze procedures for more discourse or conversational assessment. In working with adolescents, Brown, Anderson, Shillicock & Yule (1984) have suggested a number of task-based activities that tap the student's ability to use transactional language during interactions.

Academically, probes activities are widely used. The current trend toward criterion-referenced assessment in second language testing (Cloud, 1991; Cziko, 1983; Hudson & Lynch, 1984) and some applications of curriculum-based assessment (Marston & Magnusson, 1987; Tucker, 1985) are dependent on probe activities. Simon (1989) has developed a comprehensive analysis of classroom communicative abilities needed to transition from elementary to secondary school that employs several probe strategies in addition to other formats. Her work has been found to be quite successful in commenting on the success of language minority students to function in the mainstream classroom environment. When reviewed, all achievement tests and many locally-constructed measures of the academic performance of language minority students are discovered to be designed as probes (Brindley, 1986; Cloud, 1991; O'Malley, 1989) and for academic purposes, this format is very effective.

A final academic probe procedure that warrants discussion is the cloze test. As discussed by Oller (1979), and others (Hamayan, Kwait, & Perlman, 1985; Jonz, 1987) cloze procedures accurately reflect on the student's underlying language proficiency and are usually highly correlated with academic performance in English (Oller & Perkins, 1978; 1980). For example, Laesch and van Kleeck (1987) demonstrated significant correlations between their cloze procedure and the California Test of Basic Skills. The cloze procedure was effective in measuring the language needed in academic tasks and it discriminated between subjects with varying degrees of proficiency.

Behavioral Sampling

The second assessment format involves behavioral sampling procedures. Within this popular strategy, the student being assessed completes some required task and this performance is audio-recorded (or video-recorded) and transcribed or the performance is collected in some other way. After data collection, the behavioral sample is analyzed. This format has been extensively applied over the past 25 years and there are numerous procedures available for both conversational assessment and for evaluation of many academic activities.

Conversationally, a number of functional language sampling procedures are available. Loban (1976) emphasizes dimensions of clarity of expression, fluency, command of lexical expression, and comprehension. Blank and Franklin (1980) have developed an appropriateness scale to analyze spontaneous language samples and Adams and Bishop (1990) focus on exchange structures, clarity, and cohesion. The recent work of both Van Lier (1989) and Wesche (1990) show potential as functional descriptive assessment procedures for language minority students. Emphasizing the role of organizational structure, predictability, and speaker/listener rights, these researchers are attempting to focus on discourse skills that are most relevant to second language users in their second languages.

One language sample procedure that is effective in determining conversational success is Clinical Discourse Analysis (Damico, 1985a). This procedure employs a set of 17 problematic behaviors and the theoretical framework of H.P. Grice's Cooperative Principle (1975) to provide a descriptive evaluation that focuses specifically on the effectiveness, fluency, and appropriateness criteria mentioned previously. Listed below are the 17 targeted behaviors as classified within the Gricean framework:

Quantity Category
- Failure to provide significant information to the listener
- Using nonspecific vocabulary
- Informational redundancy
- Need for repetition

Quality Category
- Message inaccuracy

Relation Category
- Poor topic maintenance
- Inappropriate response
- Failure to ask relevant questions
- Situational inappropriateness
- Inappropriate speech style

Manner Category
- Linguistic non-fluency
- Revision behaviors
- Delays before responding
- Failure to structure discourse
- Turn-taking difficulty
- Gaze inefficiency
- Inappropriate intonational contour

Research has indicated that these behaviors are effective in identifying students with communicative difficulty (Damico, 1985b; 1991a; Damico & Oller, 1980; Damico, Oller, & Storey; 1984).

There are a number of effective academically-related behavior sampling procedures available for performance assessment. For example, there are many excellent narrative analyses that can be used for assessment purposes. Applebee (1978) provides a narrative organizational analysis based on developmental stages involving the production of coherent text that can be adapted for assessment while Westby (1991; Westby, Van Dongen & Maggart, 1989), Garnett (1986), Hedberg and Stoel-Gammon (1986), and Roth (1986) have demonstrated the effectiveness of several complex story grammar and narrative analyses for evaluative purposes. While it is realized that narrative development and organization are highly culturally-dependent (Heath, 1986), these behavioral analyses provide important data regarding English task expectations.

Curriculum-based assessment from a subject-based perspective (Deno, 1985; Marston & Magnusson, 1987; Tucker, 1985) and reading miscue analysis (Goodman & Goodman, 1977) frequently involve behavioral sampling in order to accomplish the actual goals of the assessment. Recently, both types of procedures have been advocated as promising informal assessments in bilingual education (Navarrete et al., 1990). Two other promising approaches along these lines are Nelson's "Curriculum-based language assessment" (1989; 1991) and Creaghead's "Classroom Script Analysis" (1991). Both of these procedures use behavior analysis to determine whether or not the student has the communicative strategies (Nelson) or the interactive scripts (Creaghead) essential to effective functioning in a classroom setting.

Finally, when targeting academically-related assessment, Portfolio Assessment must be considered. This behavioral sampling procedure is currently receiving much attention in education. Arising for evaluation purposes from the literacy and language arts fields (Flood & Lapp, 1989; Jongsma, 1989; Mathews, 1990; Valencia, 1990; Wolf, 1989), this procedure is somewhat different from many behavioral sampling procedures in that a primary "evaluation" of the artifacts placed in the student's portfolio involves generalized comparisons rather than detailed analyses. Still, this procedure is very effective in documenting the student's current performance level and his/her progress over time. If portfolio assessment is used with care and if specific evaluative procedures and processes are meshed with the current concept, then this procedure should be very effective in the academic evaluation of language minority students (Moya & O'Malley, 1990).

Rating Scales and Protocols

The third format for data collection involves rating scales and protocols. This format enables the examiner to observe the student as a communicator in the context of interest and then rate or describe that student according to a set of reliable and valid indices of communication. After the observation, the examiner completes a rating scale or protocol when the student being assessed is no longer present. Two frequent variations of this format are checklists and interview questions. Typically, procedures within this format have some sort of evaluation (e.g., numerical scale, age range, semantic differential, forced judgment of appropriate/inappropriate) for each behavior on the scale or protocol.

For conversational purposes, a number of rating scales have been developed. Damico and Oller (1985) created a functional language screening instrument, Spotting Language Problems, that is an effective rating for screening school-age individuals for communication difficulties, while Mattes (1985; Mattes & Omark, 1984) and Cheng (1987) have designed several protocols involving both verbal and nonverbal behaviors which are helpful in the descriptive assessment of Spanish and Asian LEP students. A widely known descriptive protocol, the Pragmatic Protocol (Prutting and Kirchner, 1983; 1987), focuses on a large number of language usage behaviors and requires that the evaluator rate the student's ability to use these behaviors appropriately or inappropriately. Of course, a number of the well-established evaluation procedures in ESL and EFL make use of rating scales as a basis for their evaluations (e.g., ACTFL and FSI oral interview) and more are being developed. While modifications may make these procedures more relevant, many of these procedures are currently not appropriate to the needs of the students targeted in this paper.

Academically, there are rating scales, protocols, checklists, and interview questionnaires that focus on the functional needs of the student in the classroom. For example, Ortiz (1988) has offered a questionnaire consisting of 25 questions that revolve around the evaluation of the student's educational context while Cloud (1991) has provided several questionnaires to describe home background from an academic perspective, classroom environment, and previous educational experience. In terms of checklists, O'Malley (1989) provides an "Interpersonal and Academic Skills Checklist" that focuses on 30 skills important for cognitive academic language proficiency and a "Literacy Development Checklist" to guide the evaluator in functional assessment of language minority students. In related applied linguistic fields, Nelson (1985), Creaghead and Tattershall (1985), and Larson and McKinley (1987) have also provided checklists that may be beneficial while the work of Archer and Edward (1982) and Bassett, Whittington and Staton-Spicer (1978) can be adapted for assessment within this format. Although this work was not developed originally for language minority students, these tools have been successful for our assessment purposes.

Direct and On-line Observation

The final division of assessment procedures employs direct and on-line observation. Although effective applications of this data collection approach are still relatively rare in language assessment, the approach holds promise. This format involves the direct observation of a student's communicative interaction and the real-time and immediate coding of the communicative behaviors observed. Consequently, these procedures are able to provide detailed and objective data on the speaker's performance rather than just a final judgment of sufficient/insufficient or appropriate/inappropriate communicative performance.

Two observational systems will be detailed. Both are applicable for conversational and academic evaluation. The first, Social Interactive Coding System (SICS), was designed by Rice, Sell, and Hadley (1990) to describe the speaker's verbal interactive status in conjunction with the setting, the conversational partner, and the activities in which the speaker is engaged. This tool requires a 20-minute observational period during which the evaluator observes and codes free play for 5 minutes and then takes a 5 minute break to fill in any codes which might have been missed. This "5 minutes on, 5 minutes off" format is followed for four consecutive cycles until the 20 minutes of direct observation is accomplished. This procedure was designed for use in a bilingual preschool setting but can be modified for other age groups.

The second direct observational procedure is Systematic Observation of Communicative Interaction (SOCI) (Damico, 1985b; 1991b). This tool was designed to employ a balanced set of low inference and high inference items to achieve a reliable coding of illocutionary acts, verbal and nonverbal problematic behaviors, and a determination of the appropriateness of the student's communicative interaction. Once trained to identify and code the behaviors, the evaluator observes the student for 12 minutes and codes the interactions observed each 10 seconds. This yields 72 coded cells of data per observation. The evaluator observes the student from four to seven times and this allows for sufficient data to make representative descriptions of behavior. This tool has very high reliability and validity indices (Damico, 1985b).

The Explanatory Analysis Procedure

As previously discussed, explanatory analysis involves a deeper analysis of the data collected during the descriptive analysis stage to find how/why the student exhibits the communicative difficulties documented. To answer this question, the evaluator must determine whether the problematic behaviors are due to factors extrinsic to the student or due to intrinsic difficulties at the student's deeper level of language proficiency or semiotic capacity. The true language-learning disabled student will have intrinsic explanatory factors.

While there are several ways to conduct explanatory analysis, this paper will briefly discuss the procedure reported by Damico (1991a). According to this procedure, analysis proceeds with the evaluator asking a series of questions that enable a systematic review of those variables that might have contributed to the communicative difficulties in English. Since detailed discussion is reported elsewhere (Damico, 1991a), only the questions will be provided.

In analyzing the communicative difficulties revealed during descriptive analysis, the evaluator should apply two general sets of questions. First, regarding **extrinsic explanatory factors**:

1. Are there any overt variables that immediately explain the communicative difficulties in English? Among the potential considerations:

    - Are the documented problematic behaviors occurring at a frequency level that would be considered within normal limits or in random variation?
    - Were there any procedural mistakes in the descriptive analysis phase which accounts for the problematic behaviors?
    - Is there an indication of extreme test anxiety during the observational assessment in one context but not in subsequent ones?
    - Is there significant performance inconsistency between different contexts within the targeted manifestation?
    - Is there significant performance inconsistency between different input or output modalities?

2. Is there evidence that the problematic behaviors noted in the second language can be explained according to normal second language acquisition or dialectal phenomena?

3. Is there any evidence that the problematic behaviors noted in the second language can be explained according to cross-cultural interference or related cultural phenomena?

4. Are the communicative difficulties due to a documented lack of proficiency only in the second language but not in the first?

    - Is there documented evidence of normal first language proficiency?
    - Has the student received sufficient exposure to the second language to predict better current performance?

- Does the student exhibit the same types of problematic behaviors in the first language as in the second?

5. Is there any evidence that the problematic behaviors noted in the second language can be explained according to any bias effect that was in operation before, during, or after the assessment?

    - Is the student in a subtractive bilingual environment?
    - Is the student a member of a disempowered community?
    - Are negative or lowered expectations for this student held by the student, the student's family, or the educational staff?
    - Were specific indications of bias evident in the referral, administrative, scoring, or interpretative phases of the evaluation?

If there are no extrinsic explanations for the data obtained during the descriptive analysis phase of the assessment process, then there must be a greater suspicion that the targeted student does have an intrinsic impairment. If this is the case, then the student should exhibit some underlying linguistic systematicity in both languages that can account for the majority of the behaviors noted in the descriptive analysis stage. If the communicative difficulty cannot be accounted for by asking the first five questions, then the final question aimed at **intrinsic explanatory factors** should be conducted.

6. Is there any underlying linguistic systematicity to the problematic behaviors which were noted during the descriptive analysis phase? This can be determined by completion of the following steps:

    - Ensure that no overt factors account for the problematic behaviors (first five questions),
    - Isolate the turns or utterances which contain the problematic behaviors,
    - Perform a systematic linguistic analysis on these data points looking for consistency in the appearance of problematic behaviors.

This last step means taking the utterances or productions that contained the problematic difficulties and performing a co-occurring structure analysis (Muma, 1978; Damico, 1991a). This will determine if the appearance of the problematic behaviors systematically co-occurs with an increase in linguistic complexity. There are several systematic analyses which have been found to be very effective when conducting this type of analysis of the problematic behaviors. For example, to systematically analyze from a grammatical perspective, the work of Crystal (1979; 1982) and his syntactic, phonological, and prosodic profiles are very effective as is the work of Miller and

Chapman (1983). For effective semantic analyses, Crystal's PRISM (1982), Blank, Rose, and Berlin's (1978) four levels of perceptual-language distancing, and Kamhi and Johnston's propositional complexity analysis (1991) are all practical. Other effective analysis systems that take different perspectives have been described by Halliday and Hasan (1976) and Brown and Yule (1983).

If the evaluator follows the sequence of the questions for explanatory analysis, many of the students will not need a detailed co-occurring linguistic analysis. Their problematic behaviors will be explained by extrinsic variables. If this descriptive approach is implemented, language minority students will stand less chance of being mis-identified as language impaired when they only exhibit language or cultural differences and the second set of diagnostic decisions can be appropriately determined.

Conclusion

Performance assessment of language minority students must actually target and evaluate true linguistic performance. For several reasons, this is not an easy task. First, we are still too far from a sufficient understanding of language as a semiotic and behavioral phenomenon and from a sufficient understanding of measurement theory and practice to design the ideal assessment processes and procedures. Second, linguistic and communicative assessment is a complicated process that requires effort and expertise on the part of both the test designers and the test users. Good language assessment requires the services of well-trained applied linguists and behavior analysts. Third, our assessment efforts are directed to the group of students in the schools that can least afford poor application and implementation. For many of these students, tests serve as gates and evaluators as gatekeepers to prevent them from achieving their learning potential.

Given our obligations toward language minority students, however, we must attempt to do the best that we can at the present time. We must use our expertise in a proactive manner to design assessment procedures that allow us to meet the needs of our language minority students at the same time we meet the needs of our school systems and programs. These students deserve no less. Given our current knowledge base, it is possible to conduct performance assessment in an effective and efficient manner. To do so, there must be a focus on theoretically defensible procedures and processes that generate authentic and pragmatic results. By adopting a descriptive approach to assessment that utilizes a hierarchial and synergistic construct of language proficiency, this paper has provided some reasonable suggestions and options. While some aspects of this specific process and the discussed procedures may not fit the needs of many evaluators, descriptive performance assessment as an approach allows a focus on authentic behaviors from a functional perspective with enough descriptive power to supply answers to the assessment questions of interest in the schools. By implementing a descriptive performance assessment approach, we can serve as agents of our school systems and as advocates for the language minority students that we serve and care about. As professionals, we deserve no less.

References

Adams, C. & Bishop, D.V.M. (1990). Conversational characteristics of children with semantic-pragmatic disorder. I: Exchange structure, turntaking, repairs, and cohesion. British Journal of Disorders of Communication, 24, 211-239.

Applebee, A.N. (1978). The child's concept of story. Chicago: University of Chicago.

Archer, P. and Edward, J.R. (1982). Predicting school achievement from data on pupils obtained from teachers: Towards a screening device for disadvantaged. Journal of Educational Psychology, 74, 761-770.

Bachman, L.F. (1990a). Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, L.F. (1990b). Constructing measures and measuring constructs. In B. Harley, J. Cummins, M. Swain and & P. Allen (Eds.), The development of second language proficiency (pp. 26-38), Cambridge: Cambridge University Press.

Bachman, L.F. & Palmer, A.S. (1982). The construct validation of some components of language proficiency. TESOL Quarterly, 16, 449-465.

Bassett, R.E., Whittington, N., and Staton-Spicer, A. (1978). The basics in speaking and listening for high school graduates: What should be assessed? Communication Education, 27, 300-307.

Blank, M. and Franklin, E. (1980). Dialogue with preschoolers: A cognitively-based system of assessment. Applied Psycholinguistics, 1, 127-150.

Briere, E. (1973). Cross-cultural biases in language testing. In J.W. Oller and J.C. Richards (Eds.). Focus on the learner: Pragmatic perspectives for the language teacher (pp. 214-230). Rowley, MA: Newbury House.

Brindley, G. (1986). The assessment of second language proficiency: Issues and approaches. South Australia: National Curriculum Resource Center.

Brinton, B. & Fujiki, M. (1991). Setting the context for conversational language sampling. Best Practices in School Speech Language Pathology, 2.

Brown, G., Anderson, A., Shillicock, R., and Yule, G. (1984). Teaching talk: Strategies for production and assessment. Cambridge: Cambridge University Press.

Brown, G. and Yule, G. (1983). Discourse Analysis. Cambridge: Cambridge University Press.

Canale, M. (1983). From communicative language pedagogy. In J.C. Richards and R.W. Schmidt (Eds.), Language and communication (pp. 2-28) London: Longman.

Canale, M. (1984). A communicative approach to language proficiency assessment in a minority setting. In C. Rivera Ed.), Communicative competence approaches to language proficiency assessment: Research and application (pp. 107-122). Avon, England: Multilingual Matters.

Canale, M. (1987). Language assessment: The method is the message. In D.Tannen and J.E. Alatis (Eds.), The interdependence of theory, data, and application. (Georgetown University Roundtable) (pp. 249-262). Washington, DC: Georgetown University Press

Canale, M. (1988). The measurement of communicative competence. In R.B. Kaplan et al. (Eds.), Annual review of applied linguistics, VIII, (pp. 67-84) New York: Cambridge University Press.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1, 1-47.

Carroll, J.B. (1961). Fundamental considerations in testing for English proficiency of foreign students. In Testing the English proficiency of foreign students (pp. 31-40). Washington, DC: Center for Applied Linguistics.

Carroll, J.B. (1983). Psychometric theory and language testing. In J.W. Oller, Jr. (Ed.), Issues in language testing research (pp. 80-107). Rowley, MA: Newbury House.

Chamot, A.U. & O'Malley, J.M. (1987). The Cognitive Academic Language Learning Approach (CALLA): A bridge to the mainstream. TESOL Quarterly, 21, 227-249.

Cheng, L.L. (1987b). Assessing Asian language performance. Rockville, MD: Aspen.

Cloud, N. (1991). Educational assessment. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 220-245). Austin, TX: PRO-ED.

Cochran, C. (1989). Strategies for involving LEP students in the all-English-medium classroom: A cooperative learning approach. Washington, DC: NCBE.

Crandall, J., Spanos, G., Christian, D., Simich-Dudgeon, C., & Willetts, K. (1988). Combining language and content instruction for language minority students. Wheaton, MD: NCBE.

Creaghead, N.A. (1991). What does this student know about school? Analysis of classroom scripts for interaction. Best Practices in School Speech Language Pathology, 2.

Creaghead, N.A. and Tattershall, S.S. (1985). Observation and assessment of classroom pragmatic skills. In C.S. Simon (Ed.) Communication skills and classroom success: Assessment of language-learning disabled students (pp. 105-134). San Diego: College-Hill Press.

Cronbach, L.J. (1971) Test validation. In R. Thorndike (Ed.), Educational measurement (pp. 443-507). Washington, DC: American Council on Education.

Crystal, D. (1979). Working with LARSP. New York: Elsevier.

Crystal, D. (1982). Profiling Linguistic Disability. London: Edward Arnold.

Crystal, D. (1987). Towards a 'bucket' theory of language disability: taking account of interaction between linguistic levels. Clinical Linguistics and Phonetics, 1, 7-22.

Cummins, J.P. (1983). Language proficiency and academic achievement. In J.W. Oller, Jr. (Ed.). Issues in language testing research (pp. 108-130). Rowley, MA: Newbury House.

Cummins, J.P. (1984). Bilingualism and special education: Issues in assessment and pedagogy. San Diego: College Hill Press.

Cummins, J.P. (1986). Empowering minority students: A framework for intervention. Harvard Educational Review, 56, 18-36.

Cummins, J.P. (1989). Empowering minority students. Sacramento, CA: CABE.

Cziko, G.A. (1983). Psychometric and edumetric approaches to language testing. In J.W. Oller, Jr. (Ed.). Issues in language testing research (pp. 289-307). Rowley, MA: Newbury House.

Damico, J.S. (1985a). Clinical Discourse Analysis: A functional language assessment technique. In C.S. Simon (Ed.), Communication skills and classroom success: Assessment of language-learning disabled students (pp. 165-204) San Diego: College Hill Press.

Damico, J.S. (1985b). The effectiveness of direct observation as a language assessment technique. Unpublished doctoral dissertation. University of New Mexico. Albuquerque, NM

Damico, J.S. (1991a). Descriptive assessment of communicative ability in LEP students. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 157-218). Austin, TX: PRO-ED.

Damico, J.S. (1991b). Systematic Observation of Communicative Interaction: A valid and practical descriptive assessment technique. Best Practices in School Speech-Language Pathology, 2.

Damico, J.S. & Nye, C. (1990). Collaborative issues in multicultural populations. Best Practices in School Speech-Language Pathology, 1, 127-140.

Damico, J.S. and Oller, J.W., Jr. (1980). Pragmatic versus morphological/syntactic criteria for language referrals. Language, Speech, and Hearing Services in Schools, 9, 85-94.

Damico, J.S. and Oller, J.W., Jr. (1985). Spotting Language Problems. San Diego: Los Amigos Research Associates.

Damico, J.S., Oller, J.W., Jr., and Storey, M.E. (1983). The diagnosis of language disorders in bilingual children: Pragmatic and Surface-oriented criteria. Journal of Speech and Hearing Disorders, 48, 385-394.

Damico, J.S., Secord. W.A., & Wiig, E.H. (1991). Descriptive language assessment at school: Characteristics and design. Best Practices in School Speech Language Pathology, 2.

Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Deyes,T. (1984). Towards an authentic "discourse" cloze. Applied Linguistics, 5, 128-137.

Douglas, D., and Selinker, L. (1985). Principles for language tests within the 'discourse domains' theory of interlanguage. Language Testing, 2, 205-226.

Flood, J. & Lapp, D. (1989). Reporting reading progress: A comparison portfolio for parents. The Reading Teacher, 42, 508-514.

Fradd, S.H. (1987) The changing focus of bilingual education. In S.H. Fradd & W.J. Tikinoff (Eds.), Bilingual education and bilingual special education: A guide for administrators (pp. 45-74). Austin, TX: PRO-ED.

Fuller, B. (1982). Synergetics. New York: MacMillan.

Garnett, K. (1986). Telling tales: Narratives and learning- disabled children. Topics in Language Disorders, 6, 44-56.

Goodman, K. & Goodman, Y. (1977). Learning about the psycholinguistic processes by analyzing oral reading. Harvard Educational Review, 47, 317-333.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds) Syntax and Semantics: Speech Acts, Vol. 3 (pp. 41-59). New York: Academic Press.

Halliday, M.A.K. and Hasan, R. (1976). Cohesion in English. London: Longman.

Hamayan, E.V. & Damico, J.S. (1991) Developing and using a second language. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 40-76). Austin, TX: PRO-ED.

Hamayan, E., Kwiat, J., and Perlman, R. (1985). Assessment of language minority students: A handbook for educators. Arlington Heights, IL: Illinois Resource Center.

Hamayan, E.V. & Perlman, R. (1990). Helping language minority students after they exit from bilingual/ESL programs. Washington, DC: NCBE.

Harley, B., Cummins, J., Swain, M. & Allen, P. (1990). The nature of language proficiency. In B. Harley, J. Cummins, M. Swain & P. Allen (Eds.), The development of second language proficiency (pp. 7-25), Cambridge: Cambridge University Press.

Heath, S.B. (1986). Taking a cross-cultural look at narratives. Topics in Language Disorders, 7, 84-94.

Hedberg, N.L. and Stoel-Gammon, C. (1986). Narrative analysis: Clinical procedures. Topics in Language Disorders, 7, 58-69.

Henning, G. (1987). A guide to language testing: Development-Evaluation-research. Rowley, MA: Newbury House.

Iglesias, A. (1985). Cultural conflict in the classroom: The communicatively different child. In D.N. Ripich and F.M. Spinelli (Eds.), School discourse problems (pp. 79-96). San Diego, CA: College-Hill Press.

Jongsma, K.S. (1989). Portfolio assessment. The Reading Teacher, 43, 264-265.

Jonz, J. (1987). Textual cohesion and second-language comprehension. Language Learning, 37, 409-438.

Kagan, J. (1967). On the need for relativism. American Psychologist, 22, 131-142.

Kamhi, A.G. & Johnston, J.R. (1991). Semantic assessment: Determining propositional complexity. Best Practices in School Speech Language Pathology, 2.

Laesch, K.B. & van Kleeck, A. (1987). The cloze test as an alternative measure of language proficiency of children considered for exit from Bilingual Education. Language Learning, 37, 171-189.

Larson, V.L. and McKinley, N.L. (1987). Communication assessment and intervention strategies for adolescents. Eau Claire, WI: Thinking Publications.

Loban, W. (1976). Language development: K-12. Urbana, IL: National Council of Teachers of English.

Luria, A..R. (1981). Language and cognition. New York: J. Wiley and sons.

Marston, D. & Magnusson, D. (1987). Curriculum-based measurement: An introduction. Minneapolis,MN: Minneapolis Public Schools.

Mathews, J.K. (1990). From computer management to portfolio assessments. *The Reading Teacher*, *43*, 420-421.

Mattes, L.J. (1985). *Spanish Language Assessment Procedures: A communication skills inventory*. San Diego: Los Amigos Research Associates.

Mattes, L.J. and Omark, D.R. (1984). *Speech and language assessment for the bilingual handicapped*. San Diego, CA: College-Hill Press.

Mercer, J.R. (1984). What is a racially and culturally nondiscriminatory test? A sociological and pluralistic perspective. In C.R. Reynolds and R.T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 293-356). New York: Plenum Press.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, *35*, 1012-1027.

Miller, J. and Chapman, R. (1983). *SALT: Systematic analysis of language transcripts, User's manual*. Madison: University of Wisconsin.

Moya, S.S. and O'Malley, J.M. (1990, April) A portfolio assessment model of ESL. A paper presented at International TESOL Convention, San Francisco, CA.

Muma, J.R. (1978). *The language handbook*. Englewood Cliffs: Prentice-Hall.

Navarrete, C., Wilde, J., Nelson, C., Martinez, R. & Hargett, G. (1990). *Informal assessment in educational evaluation: Implications for bilingual education programs*. Washington, DC: NCBE.

Nelson, N.W. (1985). Teacher talk and child listening — fostering a better match. In C.S. Simon (Ed.), *Communication skills and classroom success: Assessment of language-learning disabled students* (pp. 65-104). San Diego, CA: College-Hill Press.

Nelson. N.W. (1989). Curriculum-based language assessment and intervention. *Language, Speech, and Hearing Services in Schools*, *20*, 170-184.

Nelson, N.W. (1991). Targets of curriculum-based language assessment. *Best Practices in School Speech-Language Pathology*, *2*.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CN: Yale University Press.

Ogbu, J. (1978). Minority education and caste: The American system in cross-cultural perspective. New York: Academic Press.

Oller, J.W., Jr. (1972). Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. Modern Language Journal, 56, 151-158.

Oller, J. W., Jr. (1979). Language tests at school. London: Longman.

Oller, J.W., Jr. (1983). A consensus for the 80's? In J.W. Oller, Jr. (Ed.), Issues in second language testing research (pp. 351-356). Rowley, MA: Newbury House.

Oller, J.W., Jr. (1989). Conclusions toward a rational pragmatism. In J.W. Oller, Jr. (Ed.), Language and Experience (223-250). Lanham, MD: University Press of America.

Oller, J.W., Jr. & Damico, J.S. (1991). Theoretical considerations in the assessment of LEP students. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 77-110). Austin, TX: PRO-ED.

Oller, J.W., Jr. & Perkins, K. (Eds.), (1978). Language in education: Testing the tests. Rowley, MA: Newbury House.

Oller, J.W., Jr. & Perkins, K. (Eds.), (1980). Research in language testing. Rowley, MA: Newbury House.

Olshtain, E., and Blum-Kulka, S. (1985). Cross cultural pragmatics and the testing of communicative competence. Language Testing, 2, 16-30.

O'Malley, J.M. (1989). Language proficiency testing with limited English-proficient students. Georgetown University Round Table on Languages and Linguistics (pp. 235-244). Washington, D.C.: Georgetown University Press.

O'Malley, J.M. & Chamot, A.U. (1990) Learning strategies in second language acquisition. New York: Cambridge University Press.

Ortiz, A.A. (1988). Evaluating educational contexts in which language minority students are served. Bilingual Special Education Newsletter, 7, 1-4, 7.

Ortiz, A.A. and Wilkinson, C.Y. (1987). Limited English proficient and English proficient Hispanic students with communication disorders: Characteristics at initial assessment and at reevaluation. Austin TX. The University of Texas, Handicapped Minority Research Institute on Language Proficiency.

Ortiz, A.A. and Yates, J.R. (1983). Incidence of exceptionality among Hispanics: Implications for manpower planning. NABE Journal, 7, 41-54.

Prutting, C.A. and Kirchner, D.M. (1983). Applied pragmatics. In T.M. Gallagher and C.A. Prutting (Eds.), Pragmatic assessment and intervention issues in language (pp. 29-64). San Diego: College-Hill Press.

Prutting, C.A. and Kirchner, D.M. (1987). A clinical appraisal of the pragmatic aspects of language. Journal of Speech and Hearing Disorders, 52, 105-119.

Rice, M.L., Sell, M.A., and Hadley, P.A. (1990). The Social Interactive Coding System (SICS): An on-line, clinically relevant descriptive tool. Language, Speech, and Hearing Services in Schools, 21, 2-14.

Roth, F.P. (1986). Oral narrative abilities of learning-disabled students. Topics in Language Disorders, 7, 21-30.

Seliger, H.W. (1982). Testing authentic language: the problem of meaning. Language Testing, 2, 60-73.

Shohamy, E. (1983). The stability of the oral proficiency trait on the oral interview speaking test. Language Learning, 33, 527-540.

Shohamy, E. (1988). A proposed framework for testing the oral language of second/foreign language learners. Studies in Second Language Acquisition, 10, 165-180.

Shohamy, E. (1991). Discourse analysis in language testing. Annual Review of Applied Linguistics, 11, 115-131.

Shohamy, E. and Reves, T. (1982). Authentic language tests: where from and where to? Language Testing, 2, 48-59.

Shuy, R. (1981). A holistic view of language. Research in the teaching of English, 15, 101-111.

Simon, C.S. (1989). Classroom Communication Screening Procedure for Early Adolescents: A handbook for assessment and intervention. Tempe, AZ: Communi-Cog Publications.

Spolsky, B. (1973). What does it mean to know a language? Or how do you get someone to perform his competence? In J.W. Oller & J.C. Richards (Eds.) Focus on the learner (pp. 164-176). Rowley, MA: Newbury House.

Swain, M. (1985). Large-scale communicative language testing: A case study. In Y.P. Lee, et al. (Eds.), New directions in language testing (pp. 35-46). Oxford: Pergamon Press.

Tucker, J.A. (1985). Curriculum-based assessment: An introduction. Exceptional Children, 52, 199-204.

Upshur, J.A. (1973). Productive communication testing: progress report. In J.W. Oller & J.C. Richards (Eds.) Focus on the learner (pp. 177-183). Rowley, MA: Newbury House.

Valencia, S. (1990). A portfolio approach to classroom reading assessment: The whys, whats, and hows. The Reading Teacher, 43. 338-340.

Van Lier, L. (1989). Reeling, writhing, drawling, stretching and fainting in coils: Oral proficiency interviews as conversations. TESOL Quarterly, 23, 480-508.

Wesche, M. (1987). Second language performance testing: The Ontario Test of ESL as an example. Language Testing, 4, 28-47.

Westby, C.E. (1991). Narrative analysis. Best Practices in School Speech Language Pathology, 2.

Westby, C.E., Van Dongen, R., and Maggart, Z. (1989). Assessing narrative competence. Seminars in Speech and Language, 10, 63-76.

Willig, A.C. & Ortiz, A.A. (1991). The non-biased Individualized Educational Program: Linking assessment to instruction. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 281-302). Austin, TX: PRO-ED.

Wolf, D.P. (1989). Portfolio assessment: Sampling student work. Educational Leadership, 46 (7), 35-39.

Response to Jack Damico's Presentation

J. Michael O'Malley
Georgetown University

There is a growing national interest in alternative assessment as a means of determining the knowledge and skills of students in our schools. This interest stems in part from dissatisfaction with standardized tests but also originates in theoretical arguments about how children learn and what they learn. The interest is reflected in the number of professional articles on alternative assessment, in topics addressed at conferences and in workshops, in statewide testing policies, and in the national debate on the format for a national achievement test. This national interest is compatible with and has acted to advance the concerns expressed for years about standardized tests by educators of language minority students.

The Damico paper presented at this symposium, "Performance Assessment of Language Minority Students," picks up on the interests of bilingual and English as a second language (ESL) educators for appropriate uses of assessment and suggests a variety of alternative assessment procedures for identifying and placing students acquiring English. The rationale for performance or alternative assessment, according to Damico, lies in the unique characteristics of students acquiring English in schools and in theoretical arguments concerning the hierarchical and integrative nature of language proficiency in communicative contexts. Damico describes the essential features of what he refers to as "descriptive assessment," sketches out the process for conducting the assessment, and offers a classification scheme with supportive examples to illustrate varied forms of performance assessment. Additionally, he offers an approach for both descriptive and explanatory analysis of the data from performance assessment that is intended to provide a comprehensive picture of student language proficiency in education.

Damico largely succeeds in what he sets out to do although there are some minor issues that I would differ with at various points in his paper. My principal concerns are what he did not cover under the general rubric of performance assessment, some aspects of which may be more in need of attention than the topics he raises. The best way to illustrate these concerns is to retrace some of the same ground Damico covers but from a different perspective, thereby building a foundation for the areas that I think need further discussion. While discussing these topics, I will present a rationale for alternative assessment, a definition of "academic language proficiency," and draw out the implications of these for alternative assessment in schools. Following that analysis, I will return to Damico's paper for some further comments.

Rationale for Alternative Assessment

The growing interest in alternative assessment among language minority educators has been marked by an increasing number of requests for related workshops at the Georgetown University Evaluation Assistance Center (EAC)-East in 1990 and 1991. The topics covered in these workshops have included various forms of alternative assessment and portfolio development and have been presented throughout the entire Eastern region, including Puerto Rico. Educators participating in these workshops have commented on the utility of alternative assessment for classroom applications.

The rationale for practitioner requests for information on alternative assessment lies in part in dissatisfaction with standardized tests but also stems from specific instructional needs that have not been addressed in assessment. Educators are looking for assessment that will meet multiple purposes. They are looking for assessment that can be used for identification and placement, as Damico indicates, but they are also looking for assessment that will provide a continuous record of student growth. Educators need to know how students are progressing so that they can adapt instruction to student needs, communicate indicators of progress to the student or to parents, and develop a plan for assisting the student to handle academic content in English.

The need to maintain a continuous record of student progress is an important difference from the purposes of assessment that are described by Damico. Having a continuous record of student progress requires that assessment take place periodically throughout the school year and must fit within limited time constraints when teachers have other planning and instructional responsibilities to meet. Not all of the procedures suggested by Damico meet these time constraints, and some do not seem suited to maintaining a record of student progress.

Educators who have requested EAC-East workshops are also looking for assessment that reflects multiple perspectives on student language proficiency so that they can balance one form of information against another in analyzing student performance. They are especially interested in expanding on the limited perspective permitted from the use of standardized tests with language minority students, since so often the students receive low scores due to factors that are unrelated to their actual knowledge and skills. The problem with having multiple perspectives on language proficiency is that the information needs to be integrated in a systematic way.

The integration and interpretation of information from multiple assessment need considerably more attention than Damico had the opportunity to give it. What is required is a clear focus on the purposes of the assessment, the educational goals and objectives the instruments being used are designed to assess, and a procedure for interpreting each type of assessment in relation to these objectives and to each other. I will address the interpretation issue later in commenting on applications in schools.

Another benefit educators participating in these workshops hope to gain is a perspective on conducting assessment that is authentic. They are looking for assessment procedures that reflect actual tasks that students work on in classrooms rather than the relatively isolated tasks performed [illegible] responding to multiple choice tests. As Damico notes, alternative assessment can provide information that is authentic in that it reflects actual "communicative activities within real contexts" (p. 11). Damico defines authenticity in terms of linguistic realism and ecological validity. In linguistic realism, linguistic behavior is treated as a "complex and synergistic phenomenon that exists primarily for the transmission and interpretation of meaning" (p. 11). In ecological validity, communication occurs in a naturalistic setting "where true communicative performance is occurring and is influenced by contextual factors" (p. 12). The emphasis here is on the use of language for communication, a point that Damico emphasizes repeatedly, as when he notes that the interest in assessment should be on the question, "'How successful is this student as a communicator'" (p. 14).

One difficulty with Damico's approach to authenticity is that he primarily uses a linguistic rather than an academic base for language proficiency. Although he alludes to Cummins' (1984) distinction between basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP), he does not integrate the distinction into his definition of language proficiency or into the discussion of assessment instruments he describes. Furthermore, although the underlying theory on which Damico's paper is based (Oller & Damico, 1991) discusses academic language proficiency, the linguistic origin of the theory does not lead to specific recommendations for assessment of academic language skills.

A comprehensive assessment of the performance of language minority students in school will not be complete without a more precise view of the demands inherent in using academic language. Cummins' definition of academic language skills in terms of two orthogonal continua -- one focusing on the cognitive complexity of the task, and the other on the degree of contextualized support for meaning -- has served the field well in a variety of ways. Nevertheless, the definition is incomplete because it fails to describe the nature of the cognitive activity that makes academic tasks complex. The design of assessment for academic language skills needs more precision than is afforded by this general outline.

Definition of Academic Language Proficiency

My colleague Anna Chamot and I have been working for the past six years on a content based instructional method in English as a second Language that we have referred to as the Cognitive Academic Language Learning Approach (CALLA) (Chamot & O'Malley, 1987). Two distinctive features of CALLA are its incorporation of learning strategies into instructional activities and the inclusion of academic language from the content areas into ESL objectives and tasks. While relying on Cummins' definition of CALP, we realized early on that the definition had limitations precisely because the nature of the academic task requirements that lead to cognitive complexity were incompletely specified. My description of academic language in what follows is drawn from our CALLA Handbook (Chamot & O'Malley, forthcoming) and from our earlier book developing the theoretical and research base for CALLA (O'Malley & Chamot, 1990).

Academic language can be defined in terms of the vocabulary and conventions specific to any content area but, more importantly, can be understood most clearly in terms of the language functions needed for authentic academic content. Academic language functions are essential tasks that language users must be able to perform in the different content areas, and they are what makes the task simple or complex. These functions often differ from social interactive language functions. For example, one social language function is greeting or addressing another person. Sub-categories of greeting are greeting a peer, a superior, or a subordinate, and making the greeting either formal or informal. On the other hand, academic language may involve using language functions such as identifying and describing content information, explaining a process, analyzing and synthesizing concepts, justifying opinions, or evaluating knowledge.

In many classrooms academic language tends to be unidirectional: the teacher and textbook impart information and students demonstrate their comprehension by answering oral and written questions. But academic language can also be interactive. Teachers and students can discuss new concepts, share analyses, and argue about values in both teacher-student and student-student interactions. Since academic language functions such as describing and explaining can also occur with basic interpersonal interactions, it is the specific academic context that makes these functions apply to academic language proficiency.

Language functions needed in academic content include informing, classifying, comparing, justifying, persuading, synthesizing, and evaluating, as represented in Table 1. Most of these functions are required -- or should be required -- in all content areas, including mathematics, history, science, and literature. To accomplish these functions successfully with academic content requires the use of both lower and higher order thinking skills. Lower order thinking skills and less cognitively complex tasks might include recalling facts, identifying vocabulary, and giving definitions. Higher order skills involve using language to analyze, synthesize, and evaluate. There is obviously a close connection between the difficulty of the academic language task and higher order thinking.

Table 1
Academic Language Functions

Language Function	Student Uses Language to:	Examples
Seek Information	explore the environment or acquire information	Use who, what, when, where, and how to collect information
Inform	report, explain, or describe information or procedures	Retell story or content-related information in own words, tell main ideas, summarize
Analyze	separation of whole into parts	Tells parts or features of object or idea
Compare	analyze similarities and differences in objects or ideas	Indicate similarities and differences in important parts or features of objects or ideas, outline/diagram/web, indicate how A contrasts/compares with B
Classify	sort objects or ideas into groups and give reasons	Show how A is an example of B, how A is related to B, or how A and B go together but not C and D
Predict	predict implications	Predict implications from actions or from stated text
Hypothesize	hypothesize consequences	Generate hypotheses to suggest consequences from antecedents
Justify	give reasons for an action, a decision, or a point of view	Tell why A is important, why you selected A, or why you believe A
Persuade	convince another person of a point of view	Show at least two pieces of evidence or arguments in support of a position
Solve Problems	determine solution	Given stated problem, reach solution
Synthesize	combine ideas to form a new whole	Put A together with B to make C, predict or infer C from A and B, suggest a solution for a problem
Evaluate	assess the worth of an object, opinion, or decision	Select or name criteria to evaluate, prioritize a list and explain, evaluate an object or proposition, indicate reasons for agreeing or disagreeing

BEST COPY AVAILABLE

It is because of this close interrelationship that academic language skills can best be identified by describing both the language functions and the level of thinking skills needed to perform a specific task. The student's level of proficiency is then described in terms of both the functions and the thinking skills that are employed on the task. Furthermore, the functions and thinking skills demanded by a task prescribe the complexity of the language structures and the number of independent concepts that must be integrated in performing the task. Thus, the linguistic aspects of the task are prescribed by the content. This is important, because it leads to the conclusion that the evaluator needs to analyze the academic content requirements in order to understand the language requirements of instruction. In other words, assessment of academic language does not begin with an analysis of language or with language theory, but with an analysis of the academic objectives and content requirements. This is quite different from the approach advocated by Damico.

The theory underlying the approach to assessment I am suggesting originates in cognitive psychology. A substantial body of theory has emerged describing the mental processes learners use in performing complex tasks and how these processes influence learning (e.g., Anderson, 1985; Gagné, 1985; Garner, 1987; Jones & Idol, 1990). One of the conclusions from this research is that individuals use active mental processes while learning, including the learning that occurs in second language acquisition (O'Malley & Chamot, 1990). At least a portion of these mental processes entail a higher order understanding of the requirements for learning on any particular activity and an examination of prior knowledge that will assist in the new learning. One of the other conclusions is that an important component of new learning is the domain-specific knowledge that individuals bring to the task, suggesting that an analysis of the specific content demands in any domain is important for understanding performance requirements.

Implications for Alternative Assessment

There are a variety of implications for alternative assessment in combining language functions with higher order thinking skills to define academic language proficiency. In drawing these implications, I assume that students acquiring English are enrolled in a program that will incorporate at least some form of academic content such as a bilingual program or a content-based ESL program. If the special program for students acquiring English does not contain academic content, and is limited to a language focus, the student at some point will be included in content area classes and will be called upon to understand and produce academic language.

One of the first implications of the definition of academic language advocated in this paper is that the design of alternative assessment originates with an analysis of the curriculum framework for the content areas rather than with an analysis of language or a language-based syllabus. The analysis of language demands in the context provided by content area instruction produces an understanding of the language that needs to be evaluated through alternative assessment. This analysis can take place in content area texts or by analyzing the language content area teachers use in classrooms. A second implication is that alternative assessment is best thought of as a form of domain assessment that has curriculum validity for the concepts, skills, and language used in performing academic tasks. As such, alternative assessment needs to be continuous in order to reflect students' understandings of and ability to use curriculum content introduced throughout the school year. A third implication is that alternative assessment needs to reflect the complexity of concepts, skills, and language that are integrated in performing academic tasks. Because it is difficult for any single assessment approach to capture this complexity, multiple assessment needs to be used in order to gain varied perspectives on the students' academic growth. A fourth implication is that new kinds of performance instruments are needed that will assess this complexity using authentic academic tasks in which the language functions and higher order thinking skills will be evidenced by students. Because of the authentic nature of these tasks, the assessment need not take time away from teaching but should be part of the instructional process.

One of the aspects of learning that we have come to appreciate through our studies of the application of learning strategies to instruction is the importance of the processes that students use in learning concepts and skills. Learning strategies are mental or overt procedures that students use to assist their own learning. In a CALLA program, strategies are taught directly in order to ensure that students will have a satisfactory repertoire of skills for learning academic content. Because learning strategies are among the stated outcomes of instruction, they are included among the objectives and, accordingly, are assessed. Thus, in a CALLA program, alternative assessment will include assessment of learning strategies and learning processes in addition to the assessment of academic and linguistic outcomes. This is another major difference between the assessment approach expressed here and the approach suggested by Damico.

Practical Applications to Instruction

The complex and varied requirements for alternative assessment of academic language proficiency call for a straightforward approach to the interpretation of data. A strong and visible role needs to be given to portfolio development in any discussion of alternative assessment for this reason. The design of the portfolio should be focused and systematic, and the interpretation of data in the portfolio should enable users to integrate information from a variety of different sources (Moya & O'Malley, 1990). The design should consist of a five-step portfolio development process that includes the following stages:

1. Design - the statement of the purposes of the portfolio, and the selection of a committee of teachers and other staff to design the portfolio, collect the information, and review the data;

2. Focus - a statement of instructional goals that will be assessed using the portfolio information, and selection of alternative or other assessment instruments or data collection procedures to be included in the portfolio;

3. Data Collection - assignment of responsibilities for collecting the data in addition to the data collection schedule;

4. Interpretation - design of procedures for integrating the information obtained from multiple assessment, relating it to the goals, and making it useful in instruction; and

5. Evaluation - evaluation of the portfolio process, reliability of the scoring, and the portfolio's usefulness in instructional decision making for individual students or in meeting other purposes established for the portfolio.

It is in the fourth stage of the portfolio development process that the committee formed to design and use the data from the portfolio specifies the relationships among instructional objectives, the evidence and nature of student progress, and the specific instruments that do or do not support the conclusion that the student has progressed toward the objectives. Thus, the instructional use of alternative assessment is embedded in the portfolio design. Without the portfolio, the teacher is left with an unmanageable collection of alternative assessment information that is difficult to relate to the instructional intent.

Conclusion

I have presented a different view of alternative assessment from that suggested by Damico in order to highlight the way in which I believe language needs to be assessed in schools. I do not suggest that Damico's analysis is flawed, simply that there are other ways of analyzing the assessment of language minority students in schools. My major differences with Damico concerned the rationale for using alternative assessment, the definition of language proficiency, the breadth of skills that should be assessed, and the procedures for interpretation of data from alternative assessment.

Despite these differences, there are many commonalities. We concur on the importance of alternative assessment in the education of students acquiring English and in many of the characteristics that alternative assessment should possess. Among these are authenticity, functionality, validity, and both descriptive and explanatory power. Although I did not have the opportunity to elaborate on the forms of alternative assessment I advocate, I am a supporter of performance assessment, direct observation, anecdotal reports, checklists, and rating scales and many of the other types of assessment Damico describes. Let us hope that these common points and the strengths of each approach will lead to the improvement of assessment and instruction for language minority students acquiring English.

References

Anderson, J.R. (1985). Cognitive psychology and its implications. (2nd ed.). New York: Freeman.

Chamot, A.U., & O'Malley, J.M. (1987). The Cognitive Academic Language Learning Approach (CALLA). TESOL Quarterly, 21(3), 227-49.

Chamot, A.U., & O'Malley, J.M. (forthcoming). The CALLA Handbook: How to implement the Cognitive Academic Language Learning Approach. Reading, MA: Addison-Wesley.

Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy. Clevedon, U.K.: Multilingual Matters.

Gagné, E.D. (1985). The cognitive psychology of school learning. Boston, MA: Little, Brown.

Garner, R. (1987). Metacognition and reading comprehension. Norwood, NJ: Ablex.

Jones, B.F., & Idol, L. (1990) (Eds.). Dimensions of thinking and cognitive instruction. Hillsdale, NJ: Lawrence Erlbaum.

Moya, S., & O'Malley, J.M. (1990, March). Monitoring literacy development: Looking beyond basic skills. Paper presented at the annual meetings of the Teachers of English to Speakers of Other Languages, San Francisco, CA.

Oller, J.W. Jr., & Damico, J.S. (1991). Theoretical considerations in the assessment of LEP students. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 77-110). Austin, TX: Pro-ed.

193

O'Malley, J.M., & Chamot, A.U. (1990). Learning strategies in second language acquisition. Cambridge: Cambridge University Press.

Response to Jack Damico's Presentation

Cecilia J. Navarrete
University of New Mexico

Central to the evaluation of bilingual education programs, and Title VII programs in particular, are the instruments and procedures used to assess the progress of language minority students. Dr. Damico's descriptive assessment procedure of communicative abilities offers a performance-based approach to address those evaluation needs. In applying performance-based approaches such as the descriptive assessment procedure in bilingual education we must keep in mind several factors:

Factor 1: Purpose of Assessment

First and foremost is our motives or purposes for engaging in language assessment as a useful activity. In other words, why do we want to engage in language testing in the first place? Dr. Edward DeAvila argues that, from a historical bilingual education perspective, the need to assess the language proficiency of students came about as a result of the *Lau vs Nichols* Supreme Court ruling. This ruling made school districts accountable for providing an equal education to language minority students. It was followed by amendments to the Bilingual Education Act, which provided federal funding not only to assist schools in preparing language minority students to effectively participate in school but also to assess the projects' effectiveness for their participants.

As a result of these decisions, four major evaluation purposes for bilingual education programs have evolved:

(1) identification of LEP students;

(2) placement of LEP students into appropriate programs;

(3) reclassification or exit of program students;

(4) evaluation of students' progress for instruction and evaluation of the program effectiveness).

These purposes, while part and parcel of bilingual programs, at times, function independently from each other and at times are incompatible. For example, if you consider the process for identifying students, the information used most is categorical data. These are the types of data where students are identified into levels such as non-English proficient (NEP), limited English proficient (LEP), and fluent English proficient (FEP). If other data, such as the results on vocabulary, story comprehension, or writing tests, are thrown away or not used it becomes virtually impossible to do any program planning based on student needs or measure any student progress. The reason is because all one can do with categorical data is count the number of students in each group. Therefore, it is critical that, when collecting information on individual students, procedures be considered for aggregating data into a meaningful set of indices that allow for true assessment of the students' achievements.

Overall, in looking at purposes of assessment in bilingual education and assessment approaches such as Dr. Damico's, it is important to seek a balance between what is needed at the classroom, federal, and state level. The greater the gap between the purposes of bilingual education and assessment, be it alternative or traditional standardized assessments, the greater the likelihood of distorting our understanding of the relationship between language and schooling, at least as defined by the *Lau* decision -- the framework upon which bilingual education is based.

Factor 2: Issues of Validity

In thinking further about the purposes of bilingual education assessment, it is crucial to determine how we propose to validate the kind of assessments we choose -- that is, how we propose to demonstrate what we measured for some important or real sense. Alan Davis, in his recent book Principles of Language Testing (1990), emphasizes the need to assemble evidence about any test we choose. In this light, recent work by Richard Stiggins (1990), as well as by Robert Linn, Eva Baker, and Stephen Dunbar (1990) offer criteria that are consistent with traditional psychometric standards for judging the technical adequacy of performance-based measures. These are standards that should be considered before using any performance-based approach.

The major value of these criteria is that they are aimed at maximizing the validity of performance-based assessments by including design features such as clarifying the purpose of the assessment, identifying the consequences or specifying uses to be made with the results, and defining in explicit, observable terms the tasks and performance criteria to be considered in the assessment. In general, "authentic" performance assessments of students' performance on instructional tasks must: be accurate and viable; include the fundamental constructs of measurement; and demonstrate how they will contribute to the improvement of instruction and learning, especially for LEP students.

Factor 3: Practical Constraints

Other factors to consider in bilingual evaluations are the obvious practical constraints on language assessment that affect validity. The most obvious practical problem is what Melville called the need for "time, cash, and patience," which for bilingual programs means ensuring that the facilities, the materials, the personnel and so on are available in the numbers and at the times they are required. For example, an assessment that requires individual audio recordings of hundreds, even thousands of LEP students, all to be assessed within a specified time period could prove to be difficult if there is a shortage in the number of qualified and experienced data collectors. Other practical difficulties which need to be foreseen are those such as noise conditions, materials, and/or equipment. Overall, a test or assessment device should be simplified as much as possible and limit its requirements of people, time, and materials.

Another practical constraint has to do with the issue of reliability, especially as it pertains to performance-based assessment. By reliability I am referring to the consistency by which observations, judgments, and results are interpreted. As we seek effective approaches to acquire acceptable levels of reliability there is a minimum set of standards recommended by EAC-West:

(1) design clear, observable scoring criteria in order to maximize the raters' (e.g., teachers or evaluators) understanding of the performance to be evaluated;

(2) ensure training for inter- and intra-rater reliability when more then one person is involved in the scoring process;

(3) allow time to test the observation instrument and its ability to pick up the information desired;

(4) maintain objectivity in assessing student work by periodically checking the consistency of ratings given to students' work in the same area;

(5) keep consistent and continuous records of students to measure their development and learning outcomes;

(6) check judgments by using multiple measures including standardized and other performance-based assessments.

I mention the use of standardized tests as part of the multiple measures package because their are many who are uncertain whether to support or oppose its use. Many bilingual educators have criticized them for not being applicable to their student population or program aims -- and justifiably so. On the other hand, these types of tests continue to be administered annually by school districts, and they provide a ready source of achievement or linguistic data.

For those who are struggling with the decision on how best to use data that are available in their school district, I recommend reading an article by Blain Worthen and Vicki Spandel in the February 1991 issue of Educational Leadership. They address some of the most common criticisms of standardized testing and offer suggestions on how to avoid the pitfalls of over interpretation and misuse of standardized tests. Some pitfalls they address are:

(1) using a single test score to make important decisions about students;

(2) failing to supplement test scores with other information (for example, the teacher's knowledge of the students); and

(3) assuming tests measure all the content skills or behaviors or interests.

Overall, Worthen and Spandel point out that when standardized tests are used correctly, they do have value, but they provide only part of the picture and do have limitations.

Conclusion

Obviously, there is no quick fix answers to the assessment and testing dilemma of LEP students. However, there are steps we can take to make our evaluations practical, viable, and accurate. We can: (1) maintain a clear understanding of the purposes of bilingual education programs -- from the classroom level to the policy level; (2) Educate ourselves and those involved with LEP students about evaluation and assessment; (3) Carefully avoid any misuses of tests or performance-based assessments; and (4) realize that, no matter which assessment instrument we select, each will have its limitations -- none will be able to provide us with all the answers we are seeking. There are panaceas to assessment. As we look at the types of assessment offered by Dr. Damico we must keep in mind the constraints and limitations of such assessments as well as take into consideration the purposes of assessment for our LEP students.

SEA Usage of Alternative Assessment: The Connecticut Experience

Joan Boykoff Baron[1]
Connecticut State
Department of Education

This paper is about the use of alternative assessments at the state level with a focus on the Connecticut experience. The topic is a timely one. Judging from the size of audiences attending sessions on alternative assessments at national conferences and the numbers of articles appearing on performance assessment in recent educational journals, it is fair to say that there is a growing interest in this subject among state departments of education and local school districts. Current efforts in this country in states such as Arizona, California, Connecticut, Kentucky, Maryland, and Vermont are paralleled by efforts in other countries. Recent developments in Australia, Great Britain, and the Netherlands (Raizen et al., 1990) provide evidence of an international quest for new forms of assessments which simultaneously will better serve students, teachers, and policy makers. Students will be able to self-monitor their own progress; teachers will be able to make more informed decisions about their students' levels of understanding, and policy makers can have access to accountability data that more closely mirror the skills and applications valued by society.

This new interest in performance assessment stems from both a push and a pull. The push comes from the growing dissatisfaction with this nation's over-reliance on multiple-choice tests (Baron, 1990b; Shepard, 1989; Wiggins, 1989). Many find multiple-choice tests inadequate for assessing higher order thinking skills, deep understanding of content, complex problem solving, communication, and collaboration. Others suggest that they are having a deleterious effect on instruction by encouraging teachers to fragment their curriculum and teach isolated bits and pieces that do not hang together conceptually or tell a coherent story. The pull comes from the ecological and systemic validity of performance assessment (Frederiksen and Collins, 1989). Many educators believe that performance-based assessments more closely represent the kinds of activities that we want our students to be able to undertake as members of society and that practicing for the assessment improves these valued skills and understandings.

Defining Performance Assessment

Over the past decade the term performance assessment has been used to describe many different types of tasks. At the simplest level, a performance assessment can mean a short open-ended written task requiring a student to produce a few sentences. At its most complex level, it can mean a group task in which students work for several days or weeks to design, carry out and report on an investigation on a complex loosely structured problem or even on a problem selected and framed by the students. This paper, by tracing the work in Connecticut over the last decade, reflects the full range of possibilities from the use of a calculator to solve a series of mathematics tasks to a several-day science task in which a group of students work together to design, carry out, and orally report on the results of a series of experiments.

The Potential of Performance-Based Assessment for Improving Education

In this paper, I will focus on the potential of performance-based assessment to make a meaningful contribution to the education of our nation's students. I am operating from the assumption that we as a nation are not currently satisfied with what our nation's students know and can do. Recent reports from both the National Assessment of Educational Progress (NAEP) and International Comparative Assessments (ICA) have been far from reassuring. Most Americans, beginning with our president and governors, believe that we are a nation at risk and are calling for dramatic school reform. In this paper, we will explore the possibilities inherent in using performance-based assessment as one potential lever for changing a complex educational system. There are five aspects to the contribution that revitalizing student assessment can make to the school reform effort.

Clarifying Our Goals and Values

The first requirement is that, when designing performance tasks, it is critical to begin with a clear idea of what we value. In the spirit of AMERICA 2000 (U.S. Department of Education, 1991) and other systemic school reform efforts, I am making the assumption that we are starting with a blank slate and setting out to create assessments based not on what is currently being taught or what is currently in the curriculum but, rather, on what we hope that our students will know and be able to do to function effectively in society. Simply stated, we need to develop assessments based upon what should be happening rather than what is happening. Toward this end, there is strong consensus among educators in all disciplines that what we value today are students who have a deep understanding of content and can use higher order thinking skills to solve complex and often loosely structured problems. We also put a high premium on students' ability to communicate and collaborate effectively with others. These values are shared universally -- by educators in mathematics, science, the arts and humanities, as well as policy makers, representatives of the business community, and the general public.

Providing Richer Opportunities to Assess What We Value

The second contribution of performance assessment is that it can provide much richer opportunities to assess what we value. Today, based on work in cognitive psychology, task designers are striving to provide interesting real-world contexts to serve as situations for students to integrate their knowledge of content with their knowledge of processes and procedures (Brown, Collins, & Duguid, 1989; Resnick, 1988; Wertsch, 1985). This is by no means easy to accomplish because for so many years they have been kept separate. We are also attempting to incorporate communication skills into our new assessments, calling upon students to report their findings both orally and in writing. This represents a departure from past practice in which we have tended to measure communication skills separately. Finally, despite very little experience in assessing students working together in groups, we are attempting to provide rich contexts in which groups of students can fruitfully solve complex, interesting, and important problems.

Describing Quality Performance

The third contribution of performance assessment is that it permits us to develop a language for describing quality performance. When we develop the scoring guides for teachers and students to use in evaluating students' work, we are developing a multi-faceted description of quality. We are describing the dimensions or characteristics that accompany effective performance and finding examples of students' work across the full range of quality. This can be extremely enlightening for both students and teachers. Therefore, it is important that students' work be scored and interpreted by both the students and their own teachers. In this way, students learn to self-assess their own work and to reflect upon the extent to which they are becoming more effective writers, scientists, or artists. And teachers become more secure in their judgments of the quality of their students' work that has significant ramifications for their work in assessment, curriculum, and instruction.

Setting Standards

The fourth contribution is about standard setting. Using the descriptive criteria established for judging the quality of students' performance, we can set agreed-upon levels of satisfactory and outstanding work. Here, we are asking, "How much is good enough to warrant being labeled as adequate or exemplary?" Many educators today are familiar with how this is done in judging writing samples where teachers participate in short training programs in order to be able to recognize reliably the difference between a 3 and a 4 paper. Once teachers have learned what the attributes of quality work are and have had the opportunity to examine examples of students' work at various levels of quality, they can learn to apply these criteria to new student samples. Under these conditions, different scorers will make consistent (i.e., reliable) judgments about the same student's work. Our experience in Connecticut in scoring students' work on state assessments in a variety of subject areas is that teachers find this process energizing and empowering. For many of them, this represents the first time that they have a forum in which to articulate their own standards of quality. Unfortunately, most teachers today use scoring practices based upon tacit standards that are not shared with their students or their colleagues.

Changing Educational Conversations

The fifth and perhaps most important contribution of performance assessment is that it can dramatically alter the nature of the conversations taking place in classrooms and in the broader educational community. It influences the way teachers talk to students and the way teachers talk to one another. It influences the way students look at their own work and reflect upon its quality. When students internalize a definition of what quality means and can learn to recognize it, they have developed a very valuable critical ability. They can talk with their parents and their teachers about the quality of their work and take steps to acquire the knowledge and skills required to improve it.

Once the descriptive language and the standards are in place, similar conversations can occur between teachers and parents, between administrators and teachers, and between policy makers and members of the general public. In our current mania for "total test" scores and normative comparisons, we have begun to lose our grasp on what quality work means and how we might recognize it. It is argued here that through performance-based assessment, we can take steps to regain our understanding of quality and move toward its realization. Furthermore, it is essential to recognize that being able to describe quality work can assist us in both monitoring student progress and developing a richer array of indicators of school effectiveness. It means that we will be looking at multi-faceted manifestations of student achievement and aggregating judgments on richer and more integrated examples of students' work.

Similarities and Differences between Assessment and Instructional Tasks

There is a growing number of educators around the world who believe that there is little difference between an effective performance assessment task and an effective curriculum or learning task. Burstall (1990) calls the recent British assessment tasks "bits of curriculum." Wolf (1988) refers to the Arts PROPEL assessment tasks as "episodes of learning." I have called for "blurring the edges among assessment, curriculum, and instruction (Baron, 1990b). We view assessment tasks as learning opportunities which, at their best, are explicitly designed to foster students' understandings and skills while undergoing the assessment. This is particularly true when tasks are designed for groups of students to work together to both formulate and solve real-world problems. This should not be construed to mean that we recommend assessment tasks as initial exposures to the understandings and skills being assessed. Rather, assessment tasks are seen as integrative culminating tasks in which students deepen their understandings and synthesize many separate pieces of the curriculum.

Despite the similarities between assessment and instructional tasks, there are a few important differences. Specifically, in assessment tasks as compared with instructional tasks, the role of the teacher is less intrusive. Teachers should be willing to allow their students to flounder; they shouldn't feel the need to rush in to help their students when they don't know how to solve a problem. In addition, when using performance tasks as assessment, it is important to include a set of clear criteria for judging students' performance. Thus, the notion of "teaching to the test" becomes a desirable activity when the tests are seen as an integral part of the curriculum. If we succeed in defining the "shoulds" as described above, then the assessments would serve simultaneously to articulate and embody the goals and objectives of a course of study.

Overview of Performance Assessment in Connecticut

The next five sections of this paper describe Connecticut's attempts over the past decade to develop assessments which use meaningful performance tasks to determine what students know and can do. In all cases, results from the assessments were aggregated and reported to both state-level policy makers and school-based educators. Each group received data at an appropriate level of specificity. That is, teachers received data suitable for programmatic improvement and policy makers received accountability data suitable for determining how well educational programs were working. The examples come from three assessment programs -- two that are designed to sample a small percentage of students in order to generalize to the rest are Connecticut Assessment of Educational Progress (CAEP) and Connecticut Common Core of Learning program (CCL) and the third that tests every student in grades 4, 6, and 8 in order to identify what students might be in need of remedial assistance [Connecticut Mastery Testing (CMT)].

The first part of the paper describes the CAEP program which, between 1980 and 1987, used performance assessments to assess what students know and can do in art and music, business and office education, English language arts, science, foreign language, drafting, graphic arts, and small engines. Sample exercises and their scoring rubrics are presented and described.

The second part of the paper describes the CMT program which, since 1985, has included the use of calculators for mathematical problem solving in grade 8 and the use of writing samples and note-taking exercises in grades 4, 6, and 8.

The third and longest section of the paper describes the Connecticut Common Core of Learning Assessment Program in Mathematics and Science. Together, teachers and curriculum specialists from several states developed and tried out performance-based assessment tasks often lasting several days. This component of the project is composed of complex sustained tasks in which groups of students work together to design and carry out mathematical and scientific investigations. These are administered and scored by the students' own classroom teachers who participate voluntarily and receive special training. During the 1990-91 school year, a second component was added. This consists of a set of open-ended written exercises which assesses students' conceptual understandings of "big ideas" in science and mathematics. Sample tasks and scoring systems are provided from both components of the project as well as a summary of the components of effective performance tasks.

The fourth part of the paper summarizes and sets forth some of the prerequisites for the effective use of performance-based assessments to determine what students know and can do. The final section of this paper will acknowledge some of the paradoxes inherent in using performance-based assessments with students of limited English proficiency.

Performance Assessment in The Connecticut Assessment of Educational Progress (CAEP) Program between 1980 and 1987

In the 1980s, the CAEP program conducted assessments in eleven subject areas to determine how well students statewide were performing. The emphasis was on program evaluation and not on what individual students knew and were able to do. The CAEP assessment allowed us to ease into performance assessment gradually. In a low stakes testing environment, we began with short, individual on-demand exercises which were scored by external assessors who either observed the student during the task or scored students' work later at a neutral scoring site. These assessments are organized chronologically and summarized in Figure 1, which indicates what grades were tested, how long each performance task required, and when the scoring took place. In all cases, other than those in the vocational educational areas, only a small number of randomly selected students participated in the assessment.

Figure 1
Performance Testing in the Connecticut Assessment of Educational Progress Program, 1980-87

Subject	Year	Grades Tested	Performance Task	Whole Sample or Subsample)	Administration Time	When scored? (After self-administered testing or during other-administered testing)
Art	1980-81	4, 8, 11	Draw a room wall and draw a table with people around it	Subsample period	1 class	After
Music	1980-81	4, 8, 11	Sing "America" and complete a musical phrase	Subsample	A few minutes	During
Business and Office Education	1983-84					
•Accounting		12	Make journal entries and complete a payroll record	Whole	1 class period	After
Genera Office		12	Timed typing	Whole	1 class period	After

continued

Figure 1 (Continued)

Subject	Year	Grades Tested	Performance Task	Whole Sample or Subsample)	Administration Time	When scored? (After self-administered testing or during other-administered testing)
Secretary		12	Type and compose part of a letter	Whole period	1 class	After
			Take shorthand	Whole	Part of a class period	After
English Language Arts	1983-84	4, 8, 11	Write 2 essays	Subsample	1 class period	After
			Take a dictated spelling and word usage exercise	Subsample	Part of a class period	
			Revise errors in focus, organization, support and mechanics	Subsample	1 class period	After
			Take notes from a taped lecture	Subsample	Part of a class period	After
Science	1984-85	4, 8, 11	Use scientific apparatus: weigh, measure, focus microscope, etc.	Subsample	1 class period	During
		8,11	Design and conduct an experiment	Subsample	1 class period	During
Foreign Language French German Italian Spanish	1986-87	9-12	Write a letter	Whole	1 class period	After
			Speak to an interviewer	Subsample	1 class period	During
Industrial Arts and Technology Education	1986-87					
•Drafting		12	Produce a series of drawings	Subsample	3 1/4 hours	During
Graphic Arts		12	Produce a brochure	Subsample	5 1/2 hours	During
Small Engines		12	Service and repair small engines	Subsample	3 1/4 hours	During

For information, contact Joan Boykoff Baron, Connecticut State Department of Education, P.O. Box 2219, Room 342, Hartford, CT 06145. (566-3847)

BEST COPY AVAILABLE

In the sections which follow, several of these assessments will be described in greater detail, and examples of tasks will be provided.

Art and Music: 1980-81

Our first attempt at using performance assessment was facilitated by the NAEP program that had assessed art and music using performance tasks almost a decade earlier. Our CAEP assessment used four NAEP tasks and their accompanying scoring criteria and standards. In art, students were asked to make two drawings -- one of their bedroom wall and one of a table with people seated around it (Connecticut State Department of Education, 1982).

In music, students were asked to sing "America" and complete a musical phrase. The drawings were scored after the assessment was complete; the musical performances were scored during the assessment. Using performance assessment in the arts felt natural for teachers and was a comfortable starting point for our work.

Business and Office Education: 1983-84

All twelfth grade students who completed a two-year sequence in general office, secretarial, or accounting courses participated in this assessment (totalling approximately 4,000 students). In addition to a Business Knowledge multiple-choice test, the students were asked to complete a series of tasks which corresponded to the entry-level tasks that these students would be expected to perform in the workplace when they graduated from high school within a few months of the tests. The secretarial students were asked to transcribe letters from dictation and produce a letter using appropriate letter format and composition (see Table 1). The general office students took a timed typing test which was scored on both speed and accuracy (see Table 2). The accounting students were asked to make a series of journal entries which were scored on a variety of criteria related to the correctness of the balance and the titles (see Table 3). All of the papers were scored at a central scoring site by trained Connecticut Business and Office teachers. The performance standards were established by using a combination of several widely used standard-setting procedures which involved judgments by committees of experts from both the business and education communities as well as teachers' ratings of student competence (Connecticut State Department of Education, 1985).

Table 1
Results on Letter-Typing Exercise (Secretary Text)

Scoring Category	Findings
Format	
Vertical Spacing	82% satisfactory (51% excellent; 31% acceptable - could be improved)
Margins (left-right)	66% satisfactory (26% excellent, 40% acceptable)
Date/Closing (spacing, placement)	83% satisfactory
Paragraphing Format	90% satisfactory (87% excellent7 3% acceptable)
Typing	
Typing/Proofing/Correcting	37% satisfactory (18% no typographical errors. 9% errors corrected adequately)
Hyphenation	80% correct hyphenation or no hyphens used
Spacing after Punctuation	63% spacing correct throughout letter
Omissions/Alteration of Text	84% satisfactory (69% no text changes. 15% acceptable changes)
Composition	
Content	65% all information given
Readability	62% satisfactory (16% highly readable, 46% adequate readability)
Spelling, Grammar, Punctuation	19% no errors

Table 2
Timed Writing/Typing Results (General Office Test)

Gross Words **per Minute**	% of **Students**	Errors per **5 Minutes**	% of **Students**
0-18	6.3	0-3	15.8
19-28	14.0	4-7	23.8
29-38	25.1	8-14	37.6
39-48	34.5	15-21	13.5
49-58	15.9	more than 21	9.3
more than 58	4.2		

NOTE: Standards of acceptable performance were set at 39 gross words per minute and 7.5 errors per 5 minutes, as indicated by the dashed lines above.

Table 3
Results on Journalizing Performance Exercises (Accounting Test)

Task	Percent Correct	Common Errors
Entry to close Salaries Expense Account on Dec 31 in General Journal	11%	9% incorrect figures unbalanced (whether titles correct or not) 8% incorrect account titles
Entry in General Journal to record payment of payroll and payroll taxes	11%	19% incorrect account titles
Entries in appropriate journals, given cash balance, credit memo, check payment info pertaining to a particular account.		
A. Cash Balance - Cash Receipts Journal	44%	
B. Credit Memo - General Journal	7%	8% correct w/o "credit memo" explanation 9% wrong journal
C. Cash Receipt - Cash Receipts Journal	53%	8% included sales discount
Entry for Cash Payment to Creditor- Cash Payments Journal	15%	48% ignored discount
Entry for Cash Payment of Federal Taxes - Cash Payments Journal	17%	9% wrong account title

English Language Arts: 1983-84

This assessment contained multiple-choice sections in several aspects of English Language Arts including literature, listening and note-taking skills, and writing. Using a procedure called matrix sampling, different students took different parts of the assessment. However, no attempt was made to equate the different parts of the assessment because no use was to be made of the scores of individual students. Some students were asked to write two essays -- one narrative and one persuasive. Each essay was scored on more than a dozen traits ranging from the quantity and quality of supporting details to more mechanical aspects of students' writing. Others participated in a revising test in which students were asked to read and correct another student's error-laden essay. Some students were asked to provide the supporting arguments for an essay in which the beginning and end were provided. Still others were asked to provide the beginning and end of an essay in which the middle was provided. Finally, some students took a dictation test in which they heard common homonyms used in context (e.g., to, too, two; their, there, they're). Using a sample of only a few thousand students at a grade level, this assessment gave us a very thorough picture of the writing skills of Connecticut students. These understandings could not have been obtained through multiple-choice tests. Furthermore, using standards and expectations suggested by a statewide advisory committee, it gave us a very consistent picture of students' shortcomings in producing adequate supporting details in their writing as assessed in a variety of approaches (Connecticut State Department of Education, 1985).

Science: 1984-85

This assessment included a hands-on component in which pairs of students were randomly selected to accompany a specially trained external administrator to a small room in the school. There, one member of the pair was assessed on his or her ability to use various types of scientific apparatus (e.g., scales, thermometers, microscopes, balance beams, miniscus). The other student was assessed on his or her ability to design and carry out an experiment (i.e., the Survival Task) which had been developed for the Assessment Performance Unit (APU) in Great Britain. In designing and carrying out the experiment, the students were scored by an external evaluator who watched each student working alone. The evaluator looked at how carefully the student controlled for each variable and how well the results of the experiment could be trusted. (See Figure 2 for a description of the task, the scoring elements, and the data.) Using standards and expectations suggested by our advisory committee, the results were very disappointing: Whereas approximately two-thirds of the students in both grades 8 and 11 controlled for each variable individually, only one third of the students carried out an experiment whose results could be trusted (Connecticut State Department of Education, 1986). These data proved to be very valuable to us when planning for the Common Core of Learning Assessment five years later. It reinforced the importance of having students design as well as carry out investigations, something which has been getting short shrift in most science classrooms in our nation (Baron, 1990a).

Figure 2

Statement of the Problem - The Survival Task*

Imagine you are stranded on a mountainside in cold, dry, windy weather. You can choose a jacket made from one of the two fabrics in front of you. This is what you have to find out:

Which fabric would keep you warmer?

You can use any of the things in front of you. Choose whatever you need to answer the question.

You can:
- use a can instead of a person
- put warm water inside to make it more life-like
- make it a "jacket" from the material

Make a clear record of your results and conclusions so that someone else can understand what you have found out.

It would be nice to find the answer to the problem, but how you do it is important. Your answer must be a reliable one that I can trust, so please work in a careful and scientific way.

*This task was adapted from a task developed by the Assessment Performance Unit in Great Britain.

Results of the Connecticut Assessment of Education Progress in Science 1984-85

Control - Can (both size and material)

Grade 11	Grade 8	
69	64	controlled
22	21	not controlled
5	15	irrelevant considering approach
5		no response

Control - Fabric (size and fastening)

Grade 11	Grade 8	
65	64	controlled
31	34	not controlled
4	2	no response

Control - Water (initial temperature)

Grade 11	Grade 8	
75	62	controlled
16	23	not controlled
4	15	irrelevant considering approach
5		no response

Control - Water (volume)

Grade 11	Grade 8	
69	57	controlled
23	27	not controlled
4	17	irrelevant considering approach
5		no response

Control - Measurement Intervals/Temperature Drop

Grade 11	Grade 8	
63	53	controlled
28	23	not controlled
4	22	irrelevant considering approach
5	1	no response

Control - Temperature Measurements

Grade 11	Grade 8	
90	69	all measurements within 2 degrees of test administor's readings
3	5	all except one or two measurements within 2 degrees of test administrator's readings
0	5	irrelevant considering approach
7	21	no response

Control - Measurement Schedule

Grade 11	Grade 8	
65	52	permits detection of temperature change
29	43	does not permit detection of temperature change
7	5	no response

Control - Recording of Data

Grade 11	Grade 8	
65	58	data organized and recorded clearly enough to permit appropriate interpretation
30	41	data not organized and recorded clearly enough...
5	1	no response

Control - Water (initial temperature)

Grade 11	Grade 8	
75	62	controlled
16	23	not controlled
4	15	irrelevant considering approach
5		no response

Control- Conclusion

Grade 11	Grade 8	
57	51	conclusion consistent with data
12	13	conclusion not consistent with data
25	35	conclusion not possible because of design or execution
6		no response

Control - Overall Evaluation of Experiment

Grade 11	Grade 8	
39	23	design and execution such that one could "trust" conclusion
33	37	design and execution have minor problems with could create some doubt about conclusion
23	39	design and execution such that one should have no faith in the conclusion at all
5		no response

Modern Foreign Languages: French, German, Italian, Spanish: 1986-87

Our assessment in modern foreign languages consisted of items in culture, reading, listening, speaking, and writing. Communicative proficiency was highly valued by the advisory committee and it determined to develop an assessment based on the ACTFL Guidelines which represented a scale of communicative proficiency ranging from Novice to Advanced. (The quality standards were built into the ACTFL scale itself.) The reading test used authentic materials from advertisements, menus, and newspaper articles. The listening test used tape recorded conversations and weather reports. The speaking test required an oral interview lasting up to one half hour in which a specially trained Connecticut teacher who participated in a week-long ACTFL training program interviewed students one at a time. The writing assessment consisted of a letter written to a student who would be visiting next year (see Figure 3). This assessment task was specially designed to give all participating high school students (those who had completed three or more years of a modern foreign language) a chance to write something. The letter began by asking for a description of members of the student's family and the rooms in his or her house -- both of which are generally learned very early in the study of foreign language. The present tense was called for at the beginning of the letter and the past and future tenses were required later in the letter. From this one developmentally constructed task we learned a lot about the student's level of written proficiency. Students' essays were scored as Novice, Intermediate, Intermediate High, or Advanced using the scoring rubrics displayed in Figure 3. Two specially trained Connecticut foreign language teachers scored each student's essay and the level of exact agreement was over 90 percent.

Figure 3
Connecticut Assessment of Educational Progress (CAEP) - Foreign Language Writing Test

Directions: Now that your family has been accepted to host an exchange student in the INTERPALS PROGRAM, write a letter in **Spanish** welcoming the exchange student from Cordoba who is coming to live with you. The student's name is Mercedes Sanchez Aparicio.

In your letter, write about

- your family and the house in which you live
- your school and daily activities
- your interests and hobbies
- something interesting that has happened in your school or community recently

Figure 3 (Continued)

Also, ask Mercedes for any information you would like to know about her.

WRITE YOUR LETTER IN YOUR ANSWER BOOKLET.

(RUBLICS FOR SCORING)

O	Blank paper, paper entirely in English or dialectal language.
N (Novice)	Use of high-frequency words, memorized phrases and formulaic sentences on familiar topics. Little or no creativity with the language beyond memorized patterns.
I (Intermediate)	Recombinations of learned vocabulary and structures into simple sentences. Language may be inadequate to express anything other than the most elementary ideas. Choppy sentences with frequent limited vocabulary and syntactical resources. Sentences will be high end. Often reads very much like a direct translation from English.
IH (Intermediate High)	Can write creative sentences, sometimes fairly complex ones, but not consistently. Structural forms reflecting time, tense or aspect are attempted, but the result is not always successful. An ability to describe and narrate in paragraphs is emerging, but the use of basic cohesive elements indicating transition is inconsistent. Vocabulary and structural resources allow the student to paraphase at times. Papers will often read like an academic exercise.
A (Advanced)	Able to join sentences in simple discourse on familar topics. Has sufficient writing vocabulary to express self simply with some circumlocation although the language may not be idiomatic. Good control of the most frequently-used syntactic structures (e.g., common word order patterns, coordination subordination). Writing may reflect some native-language interference, but there is a sense that the student is comfortable with the target language and can go beyond the academic task.

Drafting, Graphic Arts, and Small Engines: 1986-87

High school students who had completed a two-year sequence in drafting, graphic arts or small engines participated in this assessment. Each student took a multiple-choice test of background knowledge and a sample of students in each area was selected to participate in a performance assessment. As above, in the Business and Office assessment, these tasks represented job-entry level skills that students would be expected to have obtained before being employed. In the Drafting test, students spent more than three hours drawing a series of orthographic projections; in graphic arts, students spent more than five hours producing a brochure, and in small engines, the students spent over three hours servicing and repairing a series of small engines. Every task was scored by a trained observer from business and industry who accompanied the student throughout the time and assessed the quality of the student's product and process. In the Drafting example provided in Figure 4, the quality of the product was assessed on its accuracy; its appearance (e.g., smudges, incomplete erasures, tears, and rips); its alignment of views, including correct views, including correct projection, view selection, and view position, and its completeness and correctness with attention to missing or misrepresented lines, and the size and shape. The quality of the process was judged on its technique, including the use of instruments, the fastening and problem-solving approaches, and the construction method; the layout, including view position, spacing, and projection; the lines, with attention to density, width, and character; and the geometrics, with attention to parallelism, perpendicularity, concentricity, tangencies, and angularity. This assessment represented a major step forward in articulating the scoring criteria that are often used tacitly in assessments of this type where an expert in the field holistically assesses the quality of a student's drawing. On Figure 4, for each scoring scale, there is an asterisk next to level B. Using a combination of standard-setting approaches with teachers and representatives from industry, level B was determined to be the expected level of performance for a student entering the workplace immediately after graduation from high school (Connecticut State Department of Education, 1988).

Figure 4
Drafting Job One—Orthographic Projection

Process

	A	*B	C	D	E
1. Technique (Approach)	0-2 Errors	3-4 Errors	5-6 Errors	7-8 Errors	More Than 8 Errors

Check:
_ Use of Instruments
_ Fastening Paper to Board
_ Problem-Solving. Approach
_ Construction Methods

Figure 4 (Continued)

	A	*B	C	D	E
2. Layout (Final)	0-1 Errors	2-3 Errors	4-5 Errors	6-7 Errors	More Than 7 Errors
	Check:	_ View Position _ View Spacing _ View Projection			
3. Lines	0-2 Errors	3-4 Errors	5-6 Errors	7-8 Errors	More Than 8 Errors
	Check:	_ Density _ Width _ Character (straightness, intersections, tangency, consistency)			
4. Geometrics	0-2 Errors	3-4 Errors	5-6 Errors	7-8 Errors	More Than 8 Errors
	Check:	_ Parallelism _ Perpendicularity _ Concentricity _ Tangencies _ Angularity			
5. Accuracy	95% of Lines	90% of Lines	80% of Lines	70% of Lines	<70% of Lines
6. Appearance	95% Free	90% Free	80% Free	70% Free	<70% Free
	Check:	_ Smudges _ Incomplete Erasures _ Tears/Rips			
7. Alignment of Views	Conforms to Master	1Criterion Incorrect	2 Criteria Incorrect	3 Criteria Incorrect	>3Criteria Incorrect
	Check:	_ Correct Projection _ View Selection _ View Position			
8. Completeness & Correctness	0-2 Errors	3-4 Errors	5-6 Errors	7-8 Errors	More Than 8 Errors
	Check:	_ Missing Lines _ Misrepresentation of Lines _ Construction (not true size/shape)			

* Indicates Entry-Level Job Standard

BEST COPY AVAILABLE

Performance Assessment Tasks on the Connecticut Mastery Test in Mathematics, Reading, and Language Arts, including Writing: 1985-1991.

In 1985, Connecticut moved from a proficiency test which had been taken by every student in Grade 9 to a mastery test taken by each student in Grades 4, 6, and 8. The large majority of the test uses a multiple-choice format. However, there are three performance tasks. First, as in the ninth grade test, every student produced a writing sample which was holistically scored by two specially trained Connecticut teachers at a central scoring location. If a student fell below the standard (set by the State Board of Education at a 4 on an 8-point scale), the paper would be analytically scored by a third reader on a series of four dimensions (support, focus, organization, and mechanics). Students also participated in a note-taking exercise based on the prototype developed for the CAEP program in which they took notes from a tape-recorded lecture and then used those notes later in the test to answer a series of questions. The final set of performance tasks occurs in the eighth grade mathematics assessment that contains one part on which students use calculators to solve complex multi-step problems. Because this is a higher-stakes assessment than CAEP, teachers report that they are providing more opportunities for their students than they would be providing without the assessment -- opportunities to do more writing, take notes, and use calculators. Returning to a point made earlier, if these are skills that are highly valued by society, using appropriate performance assessments can serve an important function.

The Connecticut Common Core of Learning Assessment Program in Science and Mathematics: 1990 to Present

In 1986, Connecticut's Commissioner of Education Gerald N. Tirozzi convened a blue-ribbon committee to determine what Connecticut students should know and be able to do after completing high school. The results of their deliberations are provided in Figure 5, which summarizes the attributes and attitudes, skills and competencies, and understandings and applications that they deemed appropriate. The Common Core of Learning document (Connecticut State Board of Education, 1987) was adopted by the State Board of Education.

Figure 5

Connecticut's Common Core of Learning is organized under three major headings with subheadings that reflect significant groups of skills, knowledge and attitudes:

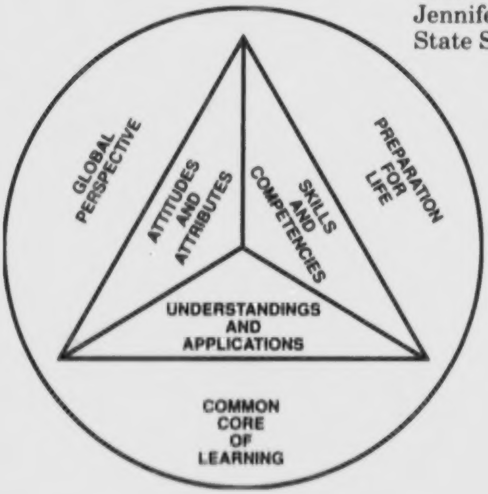


Illustration designed by
Jennifer C. Goldberg
State Student Advisory Council

Attributes and Attitudes

Self-Concept
Motivation and Persistence
Responsibility and Self-Reliance
Intellectual curiosity
Interpersonal Relations
Sense of Community
Moral and Ethical values

Skills and Competencies

Reading
Writing
Speaking, Listening
and Viewing
Quantitative Skills
Reasoning and Problem Solving
Learning Skills

Understandings and Applications

The Arts
Careers and Vocations
Cultures and Languages
History and Social Sciences
Literature
Mathematics
Physical Development and Health
Science and Technology

The Connecticut Common Core of Learning Assessment Project's overall objective is to develop performance-based assessment tasks for high school students in mathematics and science that can be used by both teachers and educational policy makers to determine what students know and can do. The content and processes included in our assessment tasks are modeled on the recommendations of mathematicians and scientists, mathematics and science educators, and representatives from business and industry. The structure of the tasks has been strongly influenced by psychological theory and research in the areas of cognition, motivation, learning and instruction. Two documents which shaped our earliest thinking in the project were The Curriculum and Evaluation Standards for School Mathematics, National Council of Teachers of Mathematics (NCTM), 1989, and Science for all Americans, (American Association for the Advancement of Science (AAAS), 1989. The first document stresses the importance of mathematics as problem solving, communication, connection making, and collaboration, and relates content to these broader purposes. The AAAS document describes the major conceptual understandings that underlie our view of the natural world as well as the appropriate attitudes and dispositions associated with science. Both documents support the view of education producing active and engaged students who are able to formulate problems, plan investigations, collect and analyze their own data, and communicate their findings effectively in writing and orally. They both envision students who are able to solve problems effectively by themselves and in groups. Connecticut's Common Core of Learning document fully supports this view of learning and assessment (Baron et al. 1989).

Some Departures from Earlier Assessment Programs.

By 1990, we felt ready to extend our performance-based assessments in several ways. First, we supplemented our on-demand tasks with embedded tasks. This approach allowed teachers to exercise choice in a number of important ways. Teachers could choose which assessment task to use and when, allowing the assessment to fit more integrally into their curriculum. Second, we extended the length of the tasks to endure over several days. Once the tasks were embedded in the classroom, it no longer mattered whether students would work at home or talk to others. Therefore, as a third departure, we included group tasks as well as individual tasks. This decision was motivated by several sources. First, there is the recognition by business and industry as well as the general public that it is important for people to be able to work as part of a team; most jobs are accomplished by a group of workers. Second, by making use of an interpersonal context, we also build upon Vygotsky's (1978) notion of the zone of proximal development. In this way, students are able to achieve a higher level of achievement earlier than they would achieve by working alone. A fourth departure resulted from our recognition of the importance of sharing the scoring criteria with students and teachers as a routine part of the assessment. This allows the kinds of conversations alluded to in the earlier part of this paper.

Three Guiding Principles.

Three additional principles have helped to shape our assessment work. The first is that we view our assessment tasks as "bits of curriculum." They are intended to provide students with opportunities to "put their learning together" -- to integrate and synthesize separate bits and pieces of knowledge about science and mathematics and deepen their understanding of the big ideas in these disciplines. The second is that we are designing our tasks to represent what our students should know rather than what they may currently be learning in their classes. This means that for the next several years, the stakes for this assessment will be low, allowing Connecticut educators time to examine their curricula, instruction, and assessment strategies in order to bring them into closer alignment with the new vision of science and mathematics. The third principle is that we view ourselves and our teachers as learners in this development process. Despite the fact that we are starting out with a fairly well articulated new vision of science and mathematics, there are few examples of consonant curriculum or assessment available. Therefore, as we deepen our own understandings of how to develop appropriate learning and assessment tasks, it is a major unfinished goal of our project staff to document and share these understandings with others.

A Description of the Common Core of Learning Assessment

Our project has two major components, both designed to provide information about what Connecticut students know and can do in science and mathematics after twelve years of school. These are described below and summarized in Figure 6.

Figure 6
Connecticut Common Core of Learning Assessment Project in Science and Mathematics: An Analysis of Its Two Components*

DIMENSION	COMPONENT I	COMPONENT II
Policy Question	Consistent with our new view of science and mathematics education, what do Connecticut high school students who are currently enrolled in science and mathematics classes know and what can they do?	Consistent with our new view of science and mathematics education, what do Connecticut high school graduates know and what can they do in science and mathematics irrespective of what courses they have taken?
Number of Tasks Pilot Tested	Mathematics: 18 Science: 26	Mathematics: 81 Science Type 1: 106 Science Type 2: 45 Science Type 3: 22
Numbers of Classrooms in which Each Task Was Administered	Between 0 and 8	Between 4 and 8
Assessment Task Format(s)/(Types)	Group investigations requiring students to design and carry out a study, analyze and portray data and report the results in writing and orally. Individual tasks precede and follow the group work.	Mathematics Open-ended problems requiring written responses, justifications and explanations. Problems have multiple solutions and/or solution paths and may require using mathematics to make decisions.
Assessment Task Format(s)/(Types)		Science: Type 1 Responding to open-ended questions and problems requiring written answers, justifications, and explanations. Type 2—Constructing charts, graphs, and tables from data and interpreting qualitative information. Type 3—Students generally design and always conduct a hands-on investigation in the presence of a trained observer who interviews the student.
Time per Task	Several class periods with some out-of-school time.	Mathematics tasks and Science Types 1 and 2: Approximately 10-20 minutes per task. Science Type 3 tasks require between one and two class periods.

Figure 6 (Continued)
Connecticut Common Core of Learning Assessment Project in Science and Mathematics: An Analysis of Its Two Components*

DIMENSION	COMPONENT I	COMPONENT II
Pilot Sample	Volunteer high school science and mathematics in 20 states administered three tasks of their choice to their own students in grades 9-12 in biology, chemistry, earth science, physics, general math, algebra, geometry, and advanced mathematics.	In 65 volunteer Connecticut high schools, science and mathematics teachers administered 6 to 9 tasks to their own students, primarily juniors. Tasks were matrix sampled so that different students took different tasks.
When administered	At each teacher's discretion spread out over the school year.	Between May 13 and May 24, 1991.
Scored Elements	Group Work (written and oral student reports); Finishing by Yourself (individual tasks).	Types 1 and 2: Open-ended written responses, graphs, tables, charts. Types 3: Hands-on investigations.
Other Available Data Sources	Beginning by Yourself (individual task); Self-assessment of behavior in groups; Videotapes of some groups working on tasks; Students' reactions to the task; Teachers' reactions to the task; Student attitude questionnaires (fall and spring) including students' self-reported grades.	Students' self-reported overall grades and grades in mathematics and science for each course taken.
Scoring Dimensions	Qualitative judgements obtained on between 4 and 10 dimensions.	Mathematics and Science Types 1 and 2: To be determined Summer and Fall 1991. (Our challenge is to capture qualitative differences within several different justifiable approaches to each question.
Scorers	The students' own science and mathematics teachers.	Connecticut science and mathematics teachers at a central location.

Figure 6 (Continued)
Connecticut Common Core of Learning Assessment Project in Science and Mathematics: An Analysis of Its Two Components*

DIMENSION	COMPONENT I	COMPONENT II
Required Professional Development	Extensive professional development and continual support of teachers in using group work, understanding the appropriate role of the teacher during the assessment, understanding important scoring procedures and exercising common standards of judgement.	Mathematics and Science Types 1 and 2: None required to administer the tasks. To score the tasks, considerable training will be required. Science Type 3: One day of training is required to administer the investigations. A second day of training is required to score students' work.
Who Will Be Assessed in 1991-92	Volunteer teachers in Connecticut and other states.	A random sample of Connecticut high school juniors.

* Funded by the Connecticut State Department of Education and the National Science Foundation (SPA-8954692) Project Director: Dr. Douglas A. Rindone, Ed.D. (203) 566-1684 Principal Investigator: Joan B. Baron, Ph.D. (203) 566-5454

Component 1 is designed to answer the policy question "Consistent with our new vision of science and mathematics, what do our high school students who are currently enrolled in science and mathematics classes know and what can they do? Our biology tasks will be administered by voluntary biology teachers to their own students during the school year; the chemistry tasks will be administered by voluntary chemistry teachers to their own students. The same will hold true for physics, earth science, and all areas of high school mathematics. It is our intent that data from these classroom-situated tasks be useful to at least three important client groups.

- Of primary importance are the students themselves. By participating in rich tasks with multidimensional scoring criteria, students will be able to monitor their own progress.

- Second, classroom teachers can use the data in assessing their students' learning and in making changes in their curriculum and instructional strategies.

- Third, these data will contribute to our reports to policy makers on the condition of education in Connecticut. While certain features of the research design are limiting (e.g., the fact that the sample is non-random limits the generalizability of the results), the richness of the data should deepen our understanding of what students know and can do in science and mathematics.

An example of a Component 1 science task called The Soda Task is provided below in Figure 7.

Figure 7
The Soda Task

Part I: Getting Started by Yourself

Name ____________________ I. D. # ______________________

You will be given two samples of soda; one regular soda containing sugar and the other one diet soda containing an artificial sweetener. Your task is to identify each sample as diet or regular based on your knowledge of physics, chemistry, and/or biology. **As in any experiment, you are not allowed to taste any of the samples.**

Come up with a list of properties of the two sodas which might help to distinguish between the samples. Write down as many as you can think of.

Written for the Connecticut State Department of Education - Sponsored by the National Science Foundation

Figure 7 (Continued)
The Soda Task

Part II: Group Work

Names____________________ I. D. #s____________________

The criteria that will be used to assess your group work are found on the Objectives Rating Form - Group. Each member of your group will also fill out the Group Performance Rating Form.

1. Make a list of **as many possible tests as your group can** think of **which** might help to distinguish between the two types of soda. Briefly explain why you think they might work. Write your answers below.

2. Now select two tests from your list to carry out. They should be the ones which your group believes would be the most effective in distinguishing between the two soda samples. Explain why you chose each of them. Show that you understand the science involved in each test.

3. Write out a complete experimental plan for each of these two tests. It should be clear enough so that someone else could easily repeat your experiments. Include a list of all the materials and equipment that you will need. Show your plan to your teacher before proceeding.

After getting approval from your teacher, carry out your experiments.

4. Record all of the results of your experiments in a clear and organized way.

5. What conclusions can be made from your experiments?

6. Make an oral presentation summarizing your experiments and results. Each member of your group should be ready to participate in any part of the presentation. Your teacher will determine the order of the presenters.

7. After hearing all the oral presentations answer the following question; if you were diabetic and had to know whether a sample of soda had sugar in it, which test would your group trust the most? Which test would your group trust the least? Explain fully why you chose each of these using complete sentences.

Written for the Connecticut State Department of Education -
Sponsored by the National Science Foundation

Figure 7 (Continued)

Part II: Objectives Rating Form - Group

Student I.D. #'s

1. ____________
2. ____________
3. ____________
4. ____________
5. ____________

Title of the Task: The Soda Task Teacher Id. # ____________

	Overall or test 1				Test 2			
The group should be able to ...	E	G	N.I.	U	E	G	N.I.	U
1. make a list of reasonable solutions to the problem.								
2. select tests based on scientific knowledge.								
	Testing against knowns (if used)				○ Adequate		○ inadequate	
3. design a controlled experiment.								
4. gather pertinent data.								
5. draw conclusions consistent with the data.	○ Yes		○ No		○ Yes		○ No	
	Discrepant Results (if reported)		○ Resolved adequately			○ Not resolved adequately		
6. select most and least effective tests based on their scientific validity.								
7. communicate the strategies and outcomes of a study through written means.								
8. collaborate effectively.								

E = Excellent G = Good N.I. = Needs Improvement U = Unacceptable

Figure 7 (Continued)

Part II: Objectives Rating Form - Oral Communication

Title of the Task: The Soda Task **Teacher Id. #** ________________ **Student Id. #** ________________

The student should be able to ...	E	G	N.I.	U
1. the content is well organized and appropriate to the task.				
2. presenters voices are clear, enthusiastic and loud enough to hear, with no distractions.				
3. presenters answer questions thoroughly and clearly.				
4. presenters maintain eye contact with the audience.				
5. visual aids are easily seen and understood.				

E = Excellent G = Good N.I. = Needs Improvement U = Unacceptable

Figure 7 (Continued)

STUDENT INSTRUCTIONS

Group Performace Rating Form
Connecticut Common Core of Learning Assessment Project

Using a Number 2 pencil, for each question, fill in the appropriate box to describe your behavior in the group during this task. Please note that items 3, 7, and 15 are different from the others; when you rate these items, be aware that you are pointing out a problem.

After you have completed your ratings, write the name of the task, its Task I.D. No. and the date below and circulate your self-ratings to each person in your group for his or her review and signature or initials. If any member of your group disagrees with your ratings of yourself, please discuss with that person the reasons for the disagreement and then decide whether or not you want to change your original rating.

Name of Task ______________________ Task I.D No. __________ Date __________

Signature or Initials of Other Group Members	Student I.D. No
1 ______________________________	______________
2 ______________________________	______________
3 ______________________________	______________
4 ______________________________	______________
5 ______________________________	______________

When each member of your group has approved and signed your rating sheet, please submit this form to your teacher.

If you cannot agree on a rating or if you wish to make comments about this process, please use the space below. Do not write your comments on the other side of this sheet.

This space may be used for COMMENTS

Thank you for participating in this project

Figure 7 (Continued)

Student Name ______________________________

Student I.D. Number ______________________________

Check One

A. GROUP PARTICIPATION	Almost Always	Often	Some times	Rarely
1. Participated in group discussion without prompting				
2. Did his or her fair share of the work				
3. Tried to dominate the group - interupted others, spoke too much				
4. Participated in the Group's Activities				
B. STAYING ON THE TOPIC	Almost Always	Often	Some times	Rarely
5. Paid attention, listened to what was being said and done				
6. Made comments aimed at getting the group back to the topic				
7. Got off the topic or changed the subject				
8. Stayed on the Topic				
C. OFFERING USEFUL IDEAS	Almost Always	Often	Some times	Rarely
9. Gave ideas and suggestions that helped the group				
10. Offered helpful criticism and comments				
11. Influenced the group's decisions and plans				
12. Offered Useful Ideas				
D. CONSIDERATION	Almost Always	Often	Some times	Rarely
13. Made positive, encouraging remarks about group members and their ideas				
14. Gave recognition and credit to others for their ideas				
15. Made inconsiderate or hostile comments about a group member.				
16. Was Considerate of Others.				
E. INVOLVING OTHERS	Almost Always	Often	Some times	Rarely
17. Got others involved by asking questions, requesting input or challenging others				
18. Tried to get the group working together to reach group agreements				
19. Seriously considered the ideas of others				
20. Involved Others.				
F. COMMUNICATING	Almost Always	Often	Some times	Rarely
21. Spoke clearly. Was easy to hear and understand				
22. Expressed ideas clearly and effectively				
23. Communicated Clearly				

BEST COPY AVAILABLE

Figure 7 (Continued)
The Soda Task

Part III: Finishing by Yourself

Name__________________ I. D. # ________________________

If you were given two samples of water, one of which is salt water and the other fresh water, which tests can you think of which might help to differentiate between the two samples. (You may use tests from the soda task or other ones.) Explain why you think each might work using complete sentences. Show that you understand the science involved.

Written for the Connecticut State Department of Education -
Sponsored by the National Science Foundation

Figure 7 (Continued)
Sample
Student Reaction Form

Date 10-4-90 Name of Perf. Task The Soda Task

If there is not enough room to answer the questions completely, please answer on the back. Thanks!

1. Did you enjoy working on this Performance Task? Explain why or why not.

I enjoyed working on this talk because it was a change from the step by step labs that we always do and I got to use my imagination

2. Describe something about this Performance Task that you liked.

I liked being able to try new and different procedures. I also liked the idea of using an everyday substance, like soda, in an experiment. It made science seem more practical.

3. Describe something about this Performance Task that you didn't like.

I didn't like having to ~~think~~ make up an experiment because I had never done anything like it before. (Now I think that making your experiment is fun.)

4. How did you feel about working in a group?

I liked working in a group. It was fun and I learned a lot about myself and others.

5. Would you like to do more group problem solving activities as a part of this class?

Yes, working in a group allows you to see problems in a different perspective. It also allows you to get a ~~[illegible]~~ better social education because you have to listen to everybody else and cooperate if you want to get any work done.

6. How do you feel about the use of performance tasks to evaluate your knowledge and skills?

I like ~~the~~ them. ~~Performance tasks~~ Taking a test can prove you remembered something, this doesn't mean you understand it. ~~the~~ In order to do a performance talk you must remember the information and understand it.

7. What, if anything, did you learn during this Performance Task?

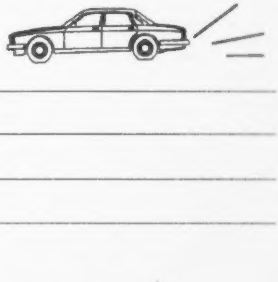
I learned to always, always label your containers and if you want something to evaporate quickly only use a small amount of the substance. I also learned that working in a group is much easier and more fun then working alone.

Component 2 is designed to answer the policy question, "What do Connecticut high school graduates know and what can they do in science and mathematics irrespective of what courses they have taken?" These performance tasks will be administered to a random sample of high school juniors by someone other than their science or mathematics teacher. (We are assessing students at the end of Grade 11 rather than Grade 12 because their motivation is higher and we believe that they will take the assessment more seriously.) Students' work will be scored by teachers at a neutral scoring site. These data will be used to report on the condition of education in Connecticut and to allow educational decision makers at all levels to set programmatic priorities for science and mathematics education. A supplemental benefit of these open-ended assessment tasks is that they will provide models of alternative formats that teachers can use to assess the depth of their students' understanding of science and mathematics. Where possible, we have attempted to write items that have several correct solutions or solution paths. Some items require students to use the same data set to support different assertions. Two examples of Component 2 tasks will be provided below in Figures 8 (science) and 9 (mathematics).

Figure 8
Energized Object

For each of the following objects, name the kinds of energy involved and explain how they are involved.

1. Moving toy car



2. Burning candle

Figure 8 (Continued)

3. Bursting balloon

4. Growing Plant

Figure 9
McDonald's Claim

You and a friend read in the newspaper that 7% of all Americans eat at McDonald's each day. Your friend says, "That's impossible!"

You know that there are approximately 250,000,000 Americans and approximately 9,000 McDonald's restaurants in the U.S. You think the claim is reasonable.

Show your mathematical work and write a paragraph or two that explains your reasoning.

Neither of these components, by itself, can provide a complete answer to the question of what our students know and can do. However, when considered together, educators and policy makers will have a better understanding of both the condition of science and mathematics education in Connecticut and some steps that can be taken to strengthen these programs.

Accomplishments to Date

During the first two years of our project, we have developed more than 300 performance tasks, described in the section which follows.

Component 1: During the 1989-90 school year, following an intensive six-day training summer session, we worked closely with fifty teachers from ten states to develop performance tasks that could be used to assess students' understandings in high school science classes (i.e., biology, chemistry, earth science, and physics) and high school mathematics classes (i.e., general mathematics, algebra, geometry, and advanced mathematics, including advanced algebra, trigonometry, and calculus). At the end of the first year, we had available approximately fifty performance tasks at different levels of development.

During July 1990, we trained a cadre of ninety high school teachers and state education department personnel to try out and refine these tasks. Before leaving the workshop, teachers were asked to choose three tasks to use in their classrooms during the 1990-91 school year. For each task, they agreed to videotape one group of their students at work, score their students' group products and processes on a series of between five and ten pre-specified scoring dimensions, and score an individual task designed to determine the extent to which each member of the group really understood what the group had done.

Each Component 1 task has three sections that involve a blend of individual work at the beginning and end of the task and group work in the middle. At the beginning of the task, each student provides information individually about his or her prior knowledge and understanding of the scientific concepts and processes relevant to the tasks. (See Figure 7, The Soda Task, Part 1 for an example.) In the middle section of the task, by far the longest phase, students work as a team to produce a group product. Students plan together and work together. Throughout the tasks, interdependence is fostered by having each student feel responsible for telling "the whole story" from the development of the group's initial design to its final conclusions. Also, at various intervals, students are asked to monitor their success both as a group and as individuals working as part of a group. (See the Checklists provided in Figure 7, Part 2 for examples of these scoring checklists.) Following the group work, a related task is administered to students individually to see what each student learned from the group experience. In the cognitive and instructional psychology literature these have been referred to as "near-transfer" or application tasks. We recognize that these individual tasks do not fully represent the knowledge tapped by the larger tasks, but they are designed to provide the teacher and students with some evidence that the student can use the knowledge gained in the group experience on a new but very similar piece of the science or mathematics terrain explored in the group task. (See Figure 7, Part 3 for an example of this near-transfer task.)

In attempting to develop a series of assessment tasks suitable for Component 1, we have developed a set of characteristics of rich performance tasks (Baron, 1990 and Baron, in press). Some of these are described in Figure 10.

Figure 10
What Are the Characteristics of Enriched Performance Assessment Tasks?

Enriched performance assessment tasks:

- are grounded in real-world contexts
- involve sustained work and often take several days of combined in-class and out-of-class time
- are based upon the most essential aspects of the content of the discipline(s) being assessed; that is. they deal with "big ideas" and major concepts (e.g., energy, form and function, change) rather than peripheral or tangential topics (American Association for the Advancement of Science, 1989; National Council of Teachers of Mathematics, 1988)
- are broad in scope, frequently integrating several scientific principles and concepts
- blend essential content with essential processes, often requiring the use of scientific methodology and the manipulation of scientific tools and apparatus
- present nonroutine, open-ended, and sometimes loosely structured problems that require students both to define the problem and to determine a strategy for solving it; optimal problems afford both multiple solutions and multiple solution paths (Charles & Saver, 1989; Greeno, 1978; Resnick, 1989; Schoenfeld, 1976)
- encourage group discussion and "brainstorming," in which a problem is considered from multiple perspectives
- require students to determine what data are needed, collect the data, report and portray them, and analyze them to discern sources of error
- call upon students to make, explain, and defend their assumptions, predictions, and estimates
- stimulate students to make connections and generalizations that will increase their understanding of the important concepts and processes

- are accompanied by explicitly stated scoring criteria related to content, process, group skills, communication skills, and a variety of motivational dispositions and "habits of mind" (Wiggins, 1989)
- spur students to monitor themselves and to think about their progress (as individuals, as members of a group, and as a complete group) in order to determine how they might improve both their investigational and group process skills
- necessitate that students use a variety of skills both for acquiring information (e.g., reading, listening, and viewing) and for communicating their strategies, data, conclusions, and reflections (e.g., speaking, writing, and graphic displays)

Baron, J. B. (1990b).

Over the past two years, we have been soliciting reactions from both the students and the teachers participating in our project. One student's reactions are found at the end of Figure 7. Although we have only begun to compile the large amount of data amassed thus far, we recognize the complexity, the difficulty, and the rewards inherent in developing meaningful and effective performance tasks. Other students' reactions were summarized by Claire Harrison (1991), a member of the CCL project team and are provided in Figure 11.

Figure 11
Student Reactions to Component 1 Tasks
Prepared by Claire Harrison
Connecticut Common Core of Learning Assessment Program

We have learned that when tasks worked well, students enjoyed the freedom and the challenge of designing and carrying out their own projects. They felt involved and intrigued, and liked not being given the answer. They liked applying and testing their knowledge, especially on a practical question. They enjoyed seeing their ideas work and their predictions confirmed, and sometimes mentioned feelings of pride and accomplishment. In order for this to occur, students needed a task that was sufficiently challenging. They also had to have an ideas of where to start and in what direction to head. Thus, they needed a level of prior knowledge about the topic. They also needed a task that was not too vague or confusing. Having a clear goal seemed important to some students.

Figure 11 (Continued)

A small minority of the students had difficulty dealing with the open-ended nature of the tasks. They were uncomfortable not knowing whether their work was correct. Some students found it helpful to be able to check their work with other group members. Whether students liked or disliked the task, most enjoyed working in a group. Working with others made the tasks more interesting and more fun. The students liked hearing the ideas and opinions of others, and finding out how others approach problems. A few mentioned enjoying having their thoughts listened to and accepted by others. Most felt they learned more by working in the group. Being able to help each other was also frequently mentioned as a positive aspect of group work. A few students did express concerns about group work. Most of these were related to the possible effect of the group on their work. They were concerned that being part of a group that worked poorly together, or in which not all members participated, would depress their own grades. Some, seeing the advantage to the group of having knowledgeable or skillful members, felt that this resource should be evenly distributed. A few students were concerned about group members who do not carry their own weight but benefit from the group's effort. A preference for working alone was expressed by a minority of students. Some of them felt they work better alone and some wanted to carry out their own ideas in their own way.

From a summary prepared by Harrison (1991) of twenty-nine teacher questionnaires returned in June 1991, we have learned:

> teachers use these assessment tasks as assessment, curriculum instruction and combinations of these. Teachers report that they are gaining important new insights about their students' skills and understandings -- expressing surprise at the difficulty their students encountered in doing the tasks. Teachers reported that they plan to use more cooperative learning and group work in their classes as a result to using these tasks. The major problem reported by the teachers involved time. Twenty-two of the twenty-nine teachers cited time as a constraint in using the tasks. Eight of these explained that the time taken to do the tasks made it difficult to cover the existing curriculum; several reported falling behind. This was a particular problem for teachers whose course of study or examinations are determined on a school-wide basis.

Component 2: The science assessment development work began in the summer of 1990 and continued throughout the fall with a selected group of Connecticut high school teachers and Department staff working together throughout the fall to write open-ended tasks. During the winter and early spring these tasks were critiqued by other Connecticut teachers and practicing scientists in Connecticut colleges and universities. In May 1991, we pilot-tested approximately 200 open-ended science items with eleventh graders in sixty Connecticut high schools. The items are composed of three types. The first consists of either a science passage to interpret or some open-ended questions to which to respond. See Figure 8 for an example of this item-type. The second item type consists of a data set to interpret. Students may be asked to construct a graph or a table and draw some conclusions from data. The third type of item is a "hands-on" experiment that students are required to design and conduct. While working, each student is observed by a trained external assessor (a retired science teacher from a different school district) who determines whether the student has designed a valid and reliable experiment and the extent to which he or she understands the relevant science content.

The mathematics tasks were developed largely by a team of mathematics educators within our department. They consisted of contextualized problems with several possible solution paths or strategies. Students were asked to communicate their reasoning to a specified audience (e.g., another student, a younger child, or an adult other than a mathematics teacher). Connecticut teachers were then convened to respond to the items and suggest improvements. During the winter and spring, the items were then reviewed by additional experts in mathematics assessment. In May 1991, we pilot-tested eighty-one open-ended items with eleventh graders in forty Connecticut high schools. (See Figure 9 for an example of a Component 2 mathematics task.)

The pilot test design provided us with between two and four classrooms of students responding to each group of items. Students responded to approximately seven tasks and also provided us with a list of courses they had taken and grades received in those courses. Teachers and students reported their reactions to the items.

Some Prerequisites for the Effective Use of Performance-Based Assessments

In reflecting on what we have learned over the past two years from listening to teachers and students participating in both components of the Common Core of Learning Assessment program, it seems obvious that new assessment approaches by themselves are insufficient. We will need to supplement new assessments with:

- Significant and sustained professional development opportunities to provide time for teachers to: identify the "big ideas" in their discipline; understand and develop a new vision of learning and teaching; develop a repertoire of new instructional strategies, and develop a sense of efficacy;

- Permissions from state and school administrators that "less is more" and that the job of teachers in David Hawkin's words is not "to cover the curriculum but to uncover the curriculum (Duchworth, 1987);

- New curriculum materials that support depth over breadth;

- Appropriate stakes and incentives so that administrators, teachers, and students will be willing to take risks and try new approaches;

- Time for teachers to develop new assessment tasks and refine them through the many iterations required;

- Time for teachers to develop shared understandings of quality and to have conversations about how to provide their students with rich opportunities to foster it;

- Time for teachers to score students' work and develop common standards.

In addition to the foregoing:

- Other high stakes tests may also need to change. We frequently hear from teachers: "We think this is the right way to teach and assess but we are too busy preparing our students to take the College Board Achievement Tests, and

- Some restructuring may be required to provide opportunities for students and teachers to achieve the higher standards we value: e.g., different configurations of class time will be required for more sustained student projects and conversations. Finally, common planning time will be necessary for teachers to work with other teachers and/or other content experts to understand what quality is and how to best achieve it.

Implications of Using Performance-Based Assessment for Students with Limited English Proficiency

Performance assessment, as developed by Connecticut, is multifaceted. It intentionally integrates content and procedural understandings with skills in problem solving, communication, and collaboration. It strives for ecological validity in determining what society values and then developing tasks which foster and assess those values most directly. This results in a strong emphasis on language skills. Students talk with one another in small groups and are called upon to communicate their findings to others at the end of their investigation. Their work rests on a foundation of content understandings. Before students can design an experiment, they have to have some knowledge about the subject of the experiment. If one uses a gate-keeper metaphor, content may serve as a gate-keeper for process, and communication skills may act as a gate-keeper for elucidating what one knows and understands. These gate keeper relationships are present for all students being assessed through the kinds of multi-faceted performance assessments advocated in this paper.

An interesting paradox surfaces in trying to build ecologically sound performance tasks. On one hand, as a society, we place high value on students being able to communicate their understandings effectively (e.g., NCTM Standards); on the other, we are concerned about the ability of minority students and students with limited English proficiency to do so. Which is more unfair -- creating high expectations for all students, while knowing that some will have difficulty, or creating relatively lower expectations for everyone, knowing that in their wake, some groups of students will not have access to demanding curricula? The answer to that question is related to the stakes imposed by the tests. If stakes are high and students are punished by poor performance on the assessments, it seems unfair to set expectations that will present hardships for certain subgroups. However, if stakes are low and better educational experiences are likely to result because of the mere existence of the assessments, then it seems unfair to deprive the groups most in need of enriched communicative experiences of those opportunities. This paradox must be addressed as states and local districts consider implementing performance-based assessments which require effective communication skills. (Linn, Baker, & Dunbar, 1991 include an interesting discussion of fairness.)

I will close as I began. Alternative assessments have grown in popularity, in part, because of the growing dissatisfaction with the fragmented and artificial multiple-choice tests that have been dominating our classrooms. Teachers have felt frustrated under the pressure to prepare their students for tests that are considered by them of dubious value. As we work toward developing alternative forms of student assessment, we must take steps to provide adequate professional development opportunities, appropriate stakes and incentives, and sufficient time and space for these innovations to take root and grow. Those of us involved in this arena of school reform believe that this endeavor is doable, difficult, and worthy of pursuit.

Notes

1. Many of the ideas in this paper resulted from my work on the Connecticut Assessment of Educational Progress Program, the Connecticut Mastery Testing Program, and the Connecticut Common Core of Learning (CCL) Assessment Program funded by the Connecticut State Department of Education. The CCL program is also funded in part by the National Science Foundation. (SPA-8954692). Many external contractors assisted the CSDE in its work and the help of these organizations is gratefully acknowledged: Advanced Systems in Measurement and Evaluation (CAEP: Science, Business and Office Education, English Language Arts), Educational Testing Service and Scholastic Testing Service (CAEP: Foreign Language), National Evaluation Systems (CAEP: Art and Music), National Occupational Competency Testing Institute (CAEP: Drafting, Graphic Arts, and Small Engines), The Psychological Corporation and Measurement Inc., (CMT). I am grateful to my colleagues at the CSDE Common Core of Learning Assessment Program for their dedicated work. The science team consists of Jeffrey Greig, Michal Lomask and Sigmund Abeles; the mathematics team consists of Bonnie Laird Hole, Susan Dixon, and Steven Leinwand. Douglas A. Rindone has provided invaluable direction for the project with the able assistance of Claire Harrison, Steven Martin and Arlene Morrissey. However, any opinions expressed in this paper are my own and are not meant to represent the views of the funding agencies, the contractors, or my coworkers.

2. In 1989, the Connecticut State Department of Education received a grant from the National Science Foundation which supported Connecticut teachers and curriculum specialists to work collaboratively with colleagues from six other states (i.e., Michigan, Minnesota, New York, Texas, Vermont, and Wisconsin) and the Coalition of Essential Schools to develop performance assessments for high school mathematics and science. After the first year, teachers from sixteen large urban school districts in the Urban Districts' Leadership Consortium of the American Federation of Teachers (including Albuquerque, NM, Cincinnati, OH, Cleveland, OH, Dade Country, FL, Detroit, MI, Hammond, IN, Kansas City, MO, Los Angeles, CA, Newark, NJ, New Orleans, LA, Philadelphia, PA, Pittsburgh, PA, Rochester, NY, Saint Paul,

MN, San Francisco, CA, and Washington, DC) and five states from Project Re:Learning (i.e., Arkansas, Delaware, New Mexico, Pennsylvania, and Rhode Island) joined the Connecticut multi-state project.

References

American Association for the Advancement of Science. (1989). Science for all Americans: A Project 2061 report on literacy goals in science, mathematics, and technology. Washington, DC: Author.

Baron, J.B. (1990a, April). How science is tested and taught in elementary school science classrooms: A study of classroom observations and interviews. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Baron, J. B. (1990b). Performance assessment: Blurring the edges among assessment, curriculum, and instruction. In A. B. Champagne, B. E. Lovitts & B. J. Calinger (Eds.) This year in school science: Assessment in the service of instruction. Washington, DC: American Association for the Advancement of Science.

Baron, J. B. (in press). Strategies for the development of effective performance exercises. In R. Stiggins and B. Plake (Eds.) Applied measurement in education.

Baron, J. B., Forgione, P. D., Rindone, D. A., Kruglanski, H., & Davey, B. (1989, April). Toward a new generation of student outcome measures: Connecticut's Common Core of Learning Assessment. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Brown, J.S., Collins, A., & Duguid, P., (1989, January-February). Situated cognition and the culture of Learning. Educational Researcher. 32-42.

Burstall, C. (1989. June; 1990, June). Update on the new national assessment in Great Britain. Paper presented at the Large Scale Assessment Conference sponsored by the Education Commission of the States and the Colorado State Department of Education, Boulder, CO.

Charles, R. I. & Silver, E. A. (Eds.). (1988). The teaching and assessing of mathematical problem solving (Vol. 3). Reston, VA: Lawrence Erlbaum Associates and the National Council of Teachers of Mathematics.

Connecticut State Board of Education (1987). Common Core of Learning. Hartford, Connecticut.

Connecticut State Department of Education (1982). *Connecticut Assessment of Educational Progress 1981-82 Art and Music Summary and Interpretations Report*. Hartford, Connecticut: Author.

Connecticut State Department of Education (1985). Connecticut Assessment of Educational Progress 1983-84. *English Language Arts Highlight Brochure*. Hartford, Connecticut.

Connecticut State Department of Education (1985). Connecticut Assessment of Educational Progress 1983-84. *Business and Office Education Summary and Interpretations Report*. Hartford, Connecticut.

Connecticut State Department of Education (1986). Connecticut Assessment of Educational Progress 1984-85. *Science Summary and Interpretations Report*. Hartford, Connecticut.

Connecticut State Department of Education (1988). Connecticut Assessment of Educational Progress 1986-87. *Drafting, Graphic Arts and Small Engines Summary and Interpretations Report and Performance Criteria Handbooks*. Hartford, Connecticut.

Duckworth, E. (1987). *The having of wonderful ideas and other essays on teaching and learning*. New York: Teachers College Press.

Frederiksen, J. R. & Collins, A. L. (1989). A systems approach to educational testing. *Educational Researcher*, *18*(9), 27-32.

Greeno, J. (1978). A study of problem solving. In R. Glaser (Ed.). *Advances in instructional psychology* (Vol. 1, pp.13-75). Hillsdale, NJ: Lawrence Erlbaum Associates.

Harrison, C. (1991) *Summary of student and teacher reactions to Component 1*. Connecticut State Department of Education (Internal documents).

Linn, R.L., Baker, E.L., & Dunbar, S. B. (1991, November) Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 15-21.

National Council of Teachers of Mathematics (1988). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Raizen, S. A., Baron, J. B., Champagne, A. B., Haertel, E., Mullis I. V. S., & Oakes, J. (1990). *Assessment in science education: The middle years* (pp. 65-75). Andover, MA: The Network.

Resnick, L.B. (1988). Learning in school and out. Educational Researcher, 16(9), 13-20.

Resnick, L.B. (1989). Teaching mathematics as an ill-structured discipline. In R. I. Charles & E. A. Silver (Eds.), The teaching and assessing of mathematical problem solving (Vol. 3, pp. 32-60). Reston, VA: Lawrence Erlbaum Associates and the National Council of Teachers of Mathematics.

Schoenfeld, A. (Ed.) (1976). Cognitive science and mathematics education. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shepard, L.A., Why we need better assessments. Educational Leadership, 46(7), 4-9.

United States Department of Education. (1991). America 2000: An education strategy. Washington, DC: Author.

Vygotsky, L.S. (1978) Mind in society.: The development of higher psychological processes. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Trans. and Eds.) Cambridge, MA: Harvard University Press. (Original work published 1935)

Wertsch, J. V. (Ed.) (1985) Culture, communication and cognition. Cambridge: Cambridge University Press.

Wiggins, G. (1989) A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 70, 703-713.

Wolf, D. (1988) Opening up assessment. Educational Leadership. 45(4), 24-29.

Response to Joan Baron's Presentation

Mary Jean Habermann
New Mexico Department of Education

Thank you, Rene, for a short introduction. Twenty minutes is a short time to talk, and five minutes of introduction takes away from this time. In relation to this particular topic, I wish to identify, first, some of the strengths of the system of performance based testing from the perspective of the practitioner. Then, I would like to briefly describe the functions of language as part of learning and outline some applicable points described in the development of alternative assessment that we are doing in New Mexico, as applied to native American languages. My final comments will center on some of the implications of performance based testing for students in programs of bilingual education.

The Connecticut Assessment of Educational Progress and the Common Core of Learning Assessment Programs contain performance-based assessment tasks for high school students. I appreciate, Dr. Joan Baron, the extensive set of materials you sent me, which provided the rationale and supportive research base for this form of testing, designed, and I quote from the materials, "to determine what students know and can do." Dr. Joan Baron likens performance assessment to "a blurring of the edges among assessment curriculum and instruction."

As a former teacher, who has dedicated all of my professional career to teaching in and through two languages, I like that definition. Having been a classroom teacher for many years and also a bilingual specialist responsible for observing bilingual instruction given throughout the state with the New Mexico Department of Education, it gives me great personal and professional pleasure to discuss performance testing from this point of view. I am not an expert in evaluation nor do I claim to have deep understandings of the technical aspects of evaluation. My comments, then, in this area will relate to the purposes of teaching, and therefore, to assessing what is taught, first, for the average English speaking child, and later, in the context of bilingual learners or for those who are becoming bilingual.

I use the latter terms in reference to these students because I know that becoming bilingual is, indeed, an expansive intellectual experience for any individual, a means by which one is able to use two linguistic and cultural systems to negotiate one's world and one's place in it. Culturally speaking, a bilingual individual is able to live, act, and participate in cultural events conducted in English and/or a language other than English, whether through literature, traditions, government, music, art, or any arena. An individual who is bilingual can be home in diverse language communities of the United States or in the countries of the world which speak the other language.

We all know that the addition of a second, third, or fourth language is expected and valued as a sign of a well-educated individual, well-cultivated individual in many countries of the world. To me, the term limited English proficient has always presented a much more limited view of the individual's intellectual and linguistic potential.

In terms of the testing process, tests can provide teachers one definitive means to ascertain whether students understand the concepts and skills being taught and also the degree to which they are learned; thus, they must be tied directly to the curriculum. Through testing, a teacher can continually reassess the teaching methods he or she utilizes and then reteach and recycle the skills and concepts needing attention.

Of the various concerns voiced by the general public regarding standardized achievement tests, one is the mismatch between achievement test results and the progress of students reported by teachers. Since teachers, however, use motivational factors and criteria to make judgments about student progress, and a paper and pencil achievement test does not and cannot, this disparity, then, will naturally exist. What these tests do provide is a measure of individual performance relative to a given set of standards. We must never forget that. It is relative to a given set of standards. These standards represent the skills and concepts deemed important for learning in the curriculum, and that curriculum represents a general American curriculum.

Annual assessment of student achievement using a standardized measure provides the teacher, the program, the district, and the state a status report, or "product" measure of a given performance at a given point of time compared to a stated expectation. With this orientation, comparability of achievement on a standardized measure can be established between the population tested and other groups nationwide, statewide, district-wide. The data produced is also useful to analyze performance trends of a given population, annually or longitudinally.

Now, a process orientation uses data generated by the measure diagnostically; that is to pin point and refine elements within the program of instruction to teach to those needs, not to teach the test...to teach to the needs. Since items tested represent items that may not have been taught, teachers have always known that test results do not necessarily represent an objective measure of what the student really knows, nor should they be interpreted as such.

It is indeed good to know that a state agency is designing and field testing an evaluation process, which ties what is taught to the items being tested. It is also tremendously important because this can foster greater accountability, as Dr. Joan Baron described, on the part of teachers. It gives the teachers responsibility for teaching, and it also gives them the tools to assess student learning using a uniform set of factors that are tied both to cognitive and affective domains.

The testing system I reviewed in the materials Dr. Baron sent to us has been designed for secondary English speakers. Once again, as a former elementary teacher in a program of bilingual education who also coordinated instruction in a secondary bilingual program, I saw how critical student involvement is for learning at all levels, for all students, regardless, of one's proficiency in the dominant language of the country. It is both valuable and valid for secondary students learning content area material. Why?

Society has changed dramatically in recent years and the demand upon the schools in preparing students to function effectively in this world have also changed. Students no longer need to just "know" facts and practice skills taught in the schools; rather, they need to know how to access information, how to evaluate it, and abstract and apply the "facts" directly to real life contexts. They need to learn how to think, how to problem solve, how to question, how to make judgments, and how to do so in a reasoned way. They need to know how to read and write, using standard grammatical forms for specific purposes, and they need to know the principles governing mathematics and science.

Students today live in a society where trends which influence them change as rapidly as they can flip the switch on the VCR, the TV, or the stereo system. This environment gives students, today, more control over their own interests. In a secondary classroom, many English speaking students seem to show difficulty attending to a lecture given about a topic unless it has immediate relationship to this instantaneous lifestyle. A teacher must almost become a magician to spark the interest of secondary students for the adult world they will enter. The type of assessment describe by Dr. Joan Baron is intended to involve secondary students then in the learning tasks while charging them with the responsibility for thinking and analyzing the material taught by the teacher.

Another aspect of the system that I find of tremendous importance is that it provides a focus on meaning. Rather then simply testing facts taught, this system tests students' ability to manipulate facts, to organize and share their knowledge, and then apply it, in highly contextualized settings.

In addition, the performance assessments prepared by the Department of Education of Connecticut appear to be very well-thought-out, based in well-founded research in testing, evaluation, and in the psychology of learning. It has also been validated through pilot testing.

I wish to commend the Connecticut Department of Education for it's leadership in this thoughtful and insightful initiative. Before I identify issues relative to this topic for bilingual learners or for those who are becoming bilingual, we must first focus our attention on the process of learning and also the relationship between learning and language, because this is a tremendously important connection.

What is learning? We talk about it all the time. There are many complex definitions, but one could say that learning takes place when the brain recognizes something in a new way. Just a little "aha"; the light goes on, so to speak. Learning is universal, and it is a unique characteristic of man resulting from his intelligence. It is, indeed, the genius of man to which we attribute the development of language because since the beginnings of time, man, a social being with intelligence, needed to communicate thoughts and ideas to others. Man's intelligence with language brought about the development of tools. These gave man leisure time for developing his artistic expression and also forms of governing, forms of educating, and forms of living as represented through the institutions within the society that evolved.

The schools represent the institution developed by man to transmit a universal body of knowledge valued by people. Now, the schools will implement a curriculum that encompasses this body of knowledge valued by society. And, in the schools, learning occurs primarily through the use of language. Whether it be Chinese, Swahili, Navaho, or English, language is the primary vehicle for learning, and students all over the world learn in and through the language they control.

For students with language and culture different from that of the schools, the desire is always the same -- that is, for the children to be successful and to accomplish learning. Bilingual, multi-cultural education recognizes that bilingual children stand to derive the same intellectual benefits that monolingual English speakers receive in the schools when instruction is given in and through their language. A well structured ESL program, part of a bilingual program, allows students to add this language to their intellectual repertoire, using methods and materials designed for second language learning.

Therefore, many of the psychological and linguistic principles of learning that apply to instruction in English will also apply to other languages. When we take this point to the point of evaluation, the same holds true. Dr. Alan Ginsburg, this morning, said testing should not discourage bilingualism.

When the purposes of evaluation is to ascertain what "students know and what students can do," the language of the child becomes a tremendously important factor. For students who are bilingual or who are becoming bilingual, both languages must be used. The language of the child should serve as the means to demonstrate mastery and understanding of the material taught. Now, if the purpose of the test is to ascertain what command of English the students have in the subject matter areas, then, the design and content of the tasks must take on a different configuration, and the results must be analyzed in terms of lexicon, syntax, semantics for second language learners. I believe, however, that the process contained in the materials, Dr. Joan Baron sent us, would probably remain the same in terms of individual work, group work, and evaluation.

I wish to discuss some ways in which this process can be modified for bilingual students or for students who are becoming bilingual by citing an example of alternatives we have recommended in the state of New Mexico. A bit of background is needed.

The state of New Mexico is perhaps the only state in the nation where several languages and cultures are part of a population mosaic which includes American Indian, Hispanic, Anglo and other ethnic groups and whose constitution has provisions for the maintenance of a bilingual citizenry. It is also the only state where the Spanish language has been used continuously since the early Spanish settlements were established after 1538. The seven languages spoken by the American Indian people are an integral part of government, religion, and aspects of daily life among each of the tribes whose elders value the use of the language in the community and generally require proficiency in it for governance. This situation has existed in New Mexico since the dawn of the Native American civilizations. It is only in very recent times that these languages have been written. In fact, for the Pueblo languages, some of the tribal governments are only now moving in this direction. The oral tradition remains as the ever-present form to transmit the values of the culture from generation to generation. One could say their "literature" exists in the oral form.

The teaching and learning of English as a second language has been both a personal as well as an institutional need for a large number of the population since the incorporation of the territory into the national framework of the United States in the mid-1800s. The schools of the state are always searching for ways and means to incorporate methods and materials which can facilitate the acquisition of English for speakers of other languages.

Up until 1986, the state testing program, designed to assess the learning needs of students in grades 3, 5, and 8, had always been done in the English language. With the passage of the Public School Reform Act of 1986, the state formulated grade level competencies for all subject matter areas. The schools of the state were charged with designing local assessment measures in each grade level to find out whether students had acquired the competencies prior to promotion to a higher level and also to provide a remediation process for those who had difficulties. For graduation, students needed to demonstrate mastery of these competencies. We restructured the state testing program to include competency based components for grades 3, 5, and 8, which accompanied the CTBS and also designed a high school competency exam. Students who did not pass this test would be given a certificate of attendance rather than a diploma.

The State Board of Education, recognizing the large numbers of students with languages other than English at their disposal for learning, provided for the development of alternatives for these students.

In order to assist districts with these new elements in the standards and help them in cases where exemption would be necessary, the New Mexico Department of Education developed a technical assistance manual entitled Recommended Procedures for Language Assessment. We also prepared state-wide training institutes for district personnel involved in evaluation and in bilingual education. For the Spanish language, we identified standardized achievement measures currently available which correlated to the content tested in the state testing program and prepared the competency exam in Spanish to meet the needs of Spanish speakers of New Mexico.

We were faced with difficulties in terms of the American Indian languages where an oral form of the test would need to be devised. We recognized that the district would need to rely upon a person who is fluent and educated in the native language to test the student's mastery of the competencies and also seek a consultant with knowledge of testing to assist in this process.

In these cases we recommended the following procedures:

1. List each competency.
2. Analyze the concepts and/or skills required in each competency.
3. Determine items and procedures within the linguistic and cultural framework of the child which correlate to each competency.
4. Determine what constitutes mastery of the competency.

5 Pilot test.

6. Administer the instrument and document the results.

For the American Indian languages, we needed to use the language and culture of the child as the means to find out his or her knowledge of general American curricular items. In some cases, translation alone would not do because of the cultures involved. This allowed the schools to find out what the child knew of the competency within his world experience.

In summary, it seems to me that in using performance tests with students who are bilingual or who are becoming bilingual, there are elements which may need to be incorporated into the process. As provided by Dr. Joan Baron, performance tests, basically, "have three parts that involve a blend of individual work in the beginning and end in group work in the middle. The work in the middle section is done as a team to produce a group product. Through a variety of accompanying assessment tools, some written (such as checklists, optional journals, logs, portfolios) and some oral and visual (i.e., video tapes of discussions and oral presentations), students have continual opportunities to provide evidence of their deepening understanding and related reflections. In order to warrant several hours of group time, tasks must meet one of two criteria: they must provide a forum in which students can work together and talk together in ways that intensify their understanding of essential scientific or mathematic concepts and processes, and/or their structure must allow students to divide a large amount of work among the group members and report their findings to the group."

Most important to this test, then, is the inner action that occurs. Since language is the key to learning, and because culture represents a group's values about the content of the curriculum encoded through language, I believe the following elements must be part of a performance based testing program for this population.

First: If, indeed, the tests are "to find out what students know and can do," then they must utilize the language of the students, so they're able to negotiate the meaning inherent in the tasks. This means that written material must be prepared in the language other than English for students who have studied in this language and, for those who have not, this means that this must be negotiated somehow, orally, through a bridging of the concepts between the two languages. It means that team work among the students, in the middle part, may have to be done bilingually, and the teachers need to understand the meanings of that if they are to fulfill the purposes of this type of assessment. In cases where the content of the task may be alien to the culture of the child, restructuring of this content will be necessary if these tasks are to be intrinsically motivating and have personal meaning. When the content of the task has no relevance, whatsoever, to the cultural framework of the child, we need to redesign those tasks so that they can build concept connections to the culture before we start teaching the general American curriculum. This is allied tremendously to meaning.

The second point I wish to make is that, since many of the principles of cooperative learning are being utilized in this plan, it would be wise to group English speakers with bilingual students. For students acquiring English, this will provide meaning-driven English language development in and among the four modalities of language (understanding, speaking, reading, and writing). Speakers of English will become sensitized to the other language and, perhaps, perk their interest in learning another language. For both groups, this will develop understanding among different ethnic groups and general appreciation for language and languages.

The third point is students must become sensitized to the fact that use of another language in the learning task does not apply lack of understanding nor the potential to understand English.

Fourth, I believe it will be necessary to design and pilot test a criteria to analyze student performance for these learners which will not penalize the student for English language manipulation that is not on par with English speakers. This criteria will need to be designed by a linguist who knows the language of the child and the semantic areas which may be affected. Finally, the English language arts component must contain tasks which assess English language performance in terms of the second language learner. To capture Dr. Jack Damico closing remarks, as tendered by Dr. Michael O'Malley earlier today, we will need to turn the research questions to target and evaluate "true linguistic performance," in terms of performance assessment. This, according to Michael O'Malley, means linguistic aspects are to be described by and through the tasks being done. Lastly is the fact that significant sustained professional development is needed for teachers implementing performance testing. This takes on a different dimension in the context of bilingual students, because we must not only provide teachers the training in what this means but also in the meanings of second language acquisition and the values of learning through two languages.

Response to Joan Baron's Presentation

Richard A. Figueroa
University of California, Davis

My apology to Dr. Joan Baron. I was asked approximately two weeks ago to change the nature of my presentation. Rather than address the issues she has raised, it was requested that I speak about California's emerging reforms in special education testing.

California's concerns about reforming the assessment process were "inspired" by a recent, federal challenge to a 1986 Injunction on the use of IQ tests with African American children. The injunction essentially broadened the 1979 *Larry P.* decision to cover not just black children being considered for Educable Mentally Retarded placement, but for any special education placement. The case is *Crawford v. Honing.*

In a hearing on this case in the U.S. Ninth Circuit Court (Spring, 1991), legal counsel for California informed the Court that the challenge to *Larry P.* (that African American children were unconstitutionally being singled out by denying them the right to an IQ test) may well be moot since the state was considering removing IQ from the diagnostic process in special education.

In the summer of 1991, Superintendent Honig's deputy, Dr. Shirley Thornton, asked me and my research team to help the state develop new policies and procedures in the area of assessment. [N.B., The statements in this article represent my own thinking on these topics and do not necessarily reflect those of the California State Department of Education]. I have gone through the whole cycle of being very pro-testing to gradually coming to realize that psychometric "diagnoses" for bilingual children, and possibly for all children, are really a needless, expensive mistake.

The rationale for removing IQ and possibly most psychometric tests in special education comes from four main findings.

The first is that now we can say, with considerable confidence, that we have found psychometric evidence of bias. The Court cases on test bias (*Larry P. v. Riles, Diana v. California State Board of Education, PASE v. Hannon, Crawford v. Honig*), since the 1970s, have drawn a lot of attention to this question. But most of the psychological community, especially the testing community, has been very successful in demonstrating that, in terms of psychometric evidence of bias, you cannot find it across ethnic groups. No matter whether you look at predictive validity, item analyses, reliabilities, or factor structures, you basically do not find evidence of psychometric bias.

Today I can report to you that we have begun to find this elusive quality of tests. We are finding it, or more accurately rediscovering it, right under our noses. In the early 1980s, Richard Duran began to alert us that Spanish language background seemed to have an impact on the predictive validities of college entrance test scores. In my own research (Figueroa, 1990) I found that IQ scores where very sensitive to bilingualism and that their predictive power declined in direct proportion to the degree of Spanish in the home. Because of the considerable implications from these data, I went back to the historical literature and found that, in fact, there is plenty of evidence sprinkled throughout the 1920s, 1930s, and 1940s showing similar outcomes for Japanese- and Chinese-speaking children. Recently, several studies have appeared with the same general findings (cited in Valdes and Figueroa, in press).

By the way, the latest edition of the Standards for Educational and Psychological Testing, for the first time, has a chapter on "Testing Linguistic Minorities." The opening statement is that for linguistic minorities, "every test given in English becomes, in part," an English language or literacy test. This is a momentous statement. It means that verbal (vocational, intellectual, achievement, personality) tests are biased when used with speakers of other languages. At the time of its publication this statement had little acknowledged, empirical support. Now, that support is more in evidence.

The second reason why IQ is under scrutiny in California is the tremendous misuse of the diagnostic process in special education. Hugh Mehan (et al, 1986) produced a superb little book titled Handicapping the Handicapped where he reports on his ethnographic study of the diagnostic process in special education in one school district in California. He found that school psychologists test until they find the "right" profile, the profile that verifies the referral for testing. He also found that school psychologists did not follow standardization procedures in testing. Poor practices in the administration and application of test scores was quite extensive. IQ anchored much of the socially constructed decisions in the "diagnoses" of learning handicaps.

The third reason for moving away form IQ is that the testing of children, particularly ethnic and bilingual children, really constitutes a form of medical malpractice. There is a group of adults known as school psychologists who have no medical training but who routinely make "diagnostic" decisions about medical conditions such as Mental Retardation, Attention Deficit Disorders, and neurological impairments (e.g., Learning Disabilities) on the basis of psychometric test scores. Some have suggested that the consequences of this professional activity are the wide national disparities in the prevalence rates for mild handicapping conditions. Some states have 3 percent of their public school population as Learning Disabled. Others have

7 percent. I would suggest that a plausible reason for such discrepancies is the practice of medicine without a license in the public schools. Some are suggesting that this Medical Model, which "looks" for the disabilities in the child and not in the curriculum, or the instruction, or system, may be just as implicated as the tests. Part of this speculation comes from the fact that even under the *Larry P. Injunction,* which proscribes the use of IQ with African American children, such children are still very over-represented in special education classes.

The final reason why psychometric tests are being reconsidered in the "diagnostic" process in California is financial. It costs the state approximately six hundred million dollars every three years to test the special education population. The unique quality of this expenditure is that it has absolutely no impact on instruction.

The reform of the special education assessment system in California begins with two initiatives. First the possible removal of IQ from all special education functions for all children in the public schools. Second, the removal of the current Medical Model which undergirds the assessment process. During the next two years, the state will undertake a multiple set of experiments aimed at determining which procedures will substitute the current assessment model and methods. The new system will be grounded on the following set of principles.

First, assessment will not focus exclusively on the child who is having problems in learning. As per the National Academy of Science's recommendation of the over-representation of ethnic children in special education, both the instructional contexts and the pupil's performance within these will be assessed.

Second, the current script which now governs testing, where an adult (often an unknown adult) presents a series of decontextualized, reductionist questions and tasks, will be changed. Rather than an unnatural communicative event where the tester cannot provide cues or feedback and where small verbal stimuli elicit small verbal and nonverbal responses, the new assessment procedures should provide for contextualized, verbally rich interactions over a long period of time.

Third, where the current methods elicit single language responses (since indeed there are no bilingual tests or norms available), the new procedures will allow for responses in L1, L2 or L1 and L2. As Valdes and Figueroa (in press) assert, the current monolingual testing practices may well be biased not just in what they do but also in what they fail to do, what they fail to account-for in bilingual pupils' mental repertoires.

Fourth, there can no longer be a single focus to assessment, such as "IQ intelligence." As constructivist frameworks point out, in mentation learners use multiple abilities for overcoming the limitations of short term memory, for using their own knowledge bases, for regulating their mental processes and for marshalling their available learning strategies. The new assessment tasks must allow for the use of multiple abilities and for the time necessary to engage them.

Fifth, "diagnosis" will no longer be a viable objective. Even if it were possible to determine who is not learning well because of subtractive bilingualism, because of poor instruction, because of the results of poverty, because of lack of schooling, because of limited English proficiency in an English-only classroom, because of a "communication handicap" or because of a "learning disability"; it makes little difference in terms of curricular or instructional needs (Rueda, 1989). A more viable objective would be the establishment of Optimal Learning Environments (Ruiz, Figueroa, Rueda and Beaumont, 1992) where pupils can "catch up" and return to the regular classrooms.

Sixth, the cadre of professionals engaged in this assessment process can no longer function as school psychologists currently do. The need is not for a testing technician. It is for an educational psychologist who is not afraid to know about curriculum and instruction; who can analyze the reading and writing process from children's work products; and who is willing to assess children in multiple contexts and in the psychopedagogical relationship described by Vygotzky.

As should be obvious by now, these reforms will extend quite beyond the area of assessment. The entire special education enterprise will be affected. It is very likely that even programs aimed at remedial interventions will also be impacted by these changes. As Ysseldyke and others have noted, children in these programs are indistinguishable from children in classes for the "mildly handicapped."

References

Figueroa, R.A. (1990). The Assessment of Linguistic Minority Group Children. In C.R. Reynolds and R. W. Kamphaus (Eds). Handbook of psychological and educational assessment of children: Volume I. intelligence and achievement. New York: Guildford Press.

Mehan, H., Hertweck, H. and Meihls, J.L. (1986). Handicapping the handicapped. Palo Alto: Stanford University Press.

Rueda, R. (1989). Defining mild disabilities with language minority students. Exceptional Children, 56, 121-129.

Ruiz, N.T., Figueroa, R. A., Rueda, R. and Beaumont, C. (1992). History and status of bilingual special education for Hispanic handicapped students. In R.V. Padilla and A.H. Benavides (Eds.) Critical Perspectives on Bilingual Education Research. Tempe, AZ: Bilingual Review Press.

Valdes, G. and Figueroa, R.A. (In press). The nature of bilingualism and the nature of testing. New York: Ablex.

Portfolio Assessment and LEP Students

Russell L. French
University of Tennessee

The Arguments For Alternative Forms of Assessment

There is both national and international demand for alternatives to present forms of student assessment. We find that demand expressed in the National Educational Goals, the products of the National Educational Goals panel and the "AMERICA 2000" strategy designed to flesh out those goals.

We also find it in publications and statements from various nationally prominent groups and in a number of state educational reform initiatives. For example, the National Governors' Association in its recent publication From Rhetoric to Action, states:

> There is considerable activity in new test development at the state and national levels by consortia of states and traditional test publishers. ...The goals are the same: creating instruments that go beyond paper-and-pencil, multiple-choice tests pegged to national norms to those that capture understanding and measure performance against high standards.[1]

In setting forth its nine-point educational agenda, Essential Components of A Successful Education System[2], the National Business Roundtable calls for a new education system which is performance- or outcome-based, and it states, "Assessment strategies must be as strong and as rich as the outcomes." The National Alliance For Restructuring Education and the National Center on Education and the Economy also demand a restructured education system that is performance-based. They state:

> A performance-based education system requires high standards and challenging goals for students, world-class curriculum and instruction that are demanding and varied, new performance assessments that measure higher-order skills, incentives for continuous improvement for students and educators, and consequences for persistent failure to improve.[3]

To understand this widespread interest in new assessments and assessment methodologies, one must understand the concerns about current testing programs. Arguments against current forms of assessment and for alternatives include:

261

1. Current standards for student performance as reflected in our tests are not high enough to meet the needs of the next century (or even today). New standards are needed. New assessments suitable for all students are needed.

2. Current tests and student evaluation procedures do not measure what all students actually know and are able to do.

3. Current standardized tests do not measure what is taught; i.e., they are not aligned with most curricula.

4. Current tests and assessment procedures do not measure adequately the higher order thinking skills and processes needed in today's and tomorrow's world, skills in which students are demonstrating weakness. Alternative, authentic assessments are needed.

5. Curriculum must be built around real life (authentic) tasks. Only real life, authentic assessments can validly and adequately assess the results of such a curriculum.

6. New assessments that can be used to compare the educational progress of school systems, schools, and individual students both nationally and internationally over time are needed.

7. To be appropriate for all students, assessments must be criterion-referenced; i.e., they must measure gains in knowledge and skills over time.

I believe most of these arguments are self-explanatory to most readers from the educational community. That is not to say that most readers agree with all of them but, together, they form the basis for the demand for new assessment technologies and the resultant activity. Later in this paper, the relevance of some of these points to assessment by portfolio, even at a class-room level, should be apparent.

Recent Developments In Performance Assessment

A significant amount of experimentation in new or refined methods of performance assessment is in progress. Much of the effort focuses on authenticity or realism of the assessments (tests), the standards against which to measure student performance, procedures for rating or scoring the new assessments, and training of educators in how to use and score them. That work is under-way in individual states and school districts and in projects of national scope. Among the well-known "national" or multi-state projects are the New Standards Project directed by Dr. Lauren Resnick (University of Pittsburgh Learning Research and Development Center) and Dr. Marc

Tucker (National Center On Education And The Economy), the Coalition of Essential Schools headed by Dr. Ted Sizer, the State Alternative Assessment Exchange initiated by the Council of Chief State School officers, and the projects (e.g., Project Zero, Project Propel) being implemented collaboratively by Harvard University and several school districts in several states.

Individual states already engaged in development of alternative assessments to replace current standardized tests include Arizona, California, Connecticut, Kentucky, Maryland and Vermont. A great many other states are contemplating restructuring of their assessment programs to include performance assessments. Among those with policy or legislation in place or near acceptance are Alabama, Colorado, Georgia, Iowa, South Carolina, and Virginia. Perhaps the most far-reaching state effort currently is that in Kentucky, where legislation mandates that an entire new assessment program consisting of performance assessments and NAEP-like tests be in place within five years. Further, the assessments created are to be appropriate for all learners in the schools, and all teachers in the state are to be trained to score the assessments and to produce similar ones for use in their own instruction. A 28 million dollar contract for this work has just been let.

Clusters of states also are discussing establishment of consortia to develop new performance assessments, both as a means of offsetting high development costs and as a means of creating assessments with meaning beyond the boundaries of a single state. It is now clear that the movement of families from place to place, especially within a geographic region, requires that sound assessment data follow the student. To appropriately place and instruct students in restructured curricula and schools; administrators and teachers must know what each individual knows and is able to do. This "cluster" activity is supported by the urging of President Bush, Secretary of Education Alexander, the National Goals Panel and the National Council on Standards and Testing to create new American Achievement Tests "capable of comparing the performance of students both nationally and internationally." It is thought that these American Achievement Tests should not be a single set of assessments developed at a national level but sets of assessments developed by clusters of states with similar curriculum frameworks and educational situations. "Cluster assessments" can then be equated to each other to provide national norms.

Three key emphases in these state, regional and national initiatives should be noted by classroom teachers and those responsible for instructing and measuring the progress of LEP students. First, there is great concern that new assessments be valid and appropriate for all students, regardless of handicap or language. Second, it is understood that if performance assessments are to replace current standardized tests, the methodologies used in those assessments must also be used in ongoing instruction. As has always been true, assessments must be aligned with what is taught and how it is taught if assessment results are to be valid. Third, there is concern that assessment be a part of instruction not apart from instruction. Therefore, there is emphasis on training teachers and administrators to develop and use performance assessments.

Observations Regarding LEP Students and Assessment

I do not pretend to be an expert in the education of limited and/or non-English proficient students. Indeed, my personal experience with these students is extremely limited. However, careful reading of recent literature on LEP learners and their instruction, discussions with persons responsible for teaching these students, experience in developing performance assessment instruments for both educators and students, recent experience with the RJR Nabisco Foundation's Next Century Schools (many have LEP students), teaching experience with at risk, K-12 learners, and some degree of common sense combine to lead me to several points for consideration by those who must assess the academic progress and ultimate achievements of LEP students.

1. There is obviously a need to assess what LEP students really know and are able to do. At issue in any assessment are its validity and reliability. In their simplest form, these concepts represent two questions: "How do I know that what I am measuring is what I really wanted to measure? (validity) How do I know that I am measuring consistently? (reliability)? Those issues are no less important to classroom assessments developed by teachers than they are to standardized tests. Experience in standardized testing has taught us that the language skills of the test taker influence his or her performance on the test, even when that test taker is supposedly English proficient. When a test is influenced in that way, the test is invalid for that particular learner. The invalidity stems from the fact that the measurement becomes a measurement of language rather than a measurement of whatever else we wanted to measure.

2. There appears to be a need to reinforce a student's native language, not destroy it. Several recent articles and papers on the instruction of LEP students report that the LEP student's self-concept, family relationships, and academic achievement suffer when instruction attempts to make him/her monolingual in English rather than bilingual or multi-lingual. Common sense also should tell us that we need an increasing number of persons pro-

ficient in two or more languages in our society to meet the increasing demands for international interaction. Why should we deplete or destroy some of our best resources?

If, then, we attempt to reinforce a student's native language in our instruction, we cannot do less in our assessments. Evaluation which allows only for the use of the English language sends a message quite contradictory to that being portrayed through instruction, and the "louder" message will be that sent through assessment. Whether we like it or not, assessment drives curriculum, or, more specifically, assessment drives students' perceptions of what is important in the curriculum. Further, assessment procedures inconsistent with instructional procedures also create an invalid test.

3. Learning styles and nonverbal communication patterns are critical to both instruction and assessment. This writer has researched the roles of both learning styles and nonverbal communication in the classroom for more than twenty years. For purposes of this paper, suffice it to say that there are at least seven different perceptual learning styles, to say nothing of varying cognitive, emotional and social styles. We know that there are learners who are print-oriented (dependent on reading), aural (dependent on listening), interactive (dependent on talking/verbalizing), visual (dependent on pictorial representations), haptic (dependent on touch and feel), kinesthetic (dependent on movement) and olfactory (dependent on smell and taste).[4] Further, greater numbers of certain types of learners are found in some cultures and backgrounds than in others.

Much also has been written about the importance of nonverbal communication in language and culture. More than 70 percent (perhaps as much as 90 percent) of whatever is communicated is communicated nonverbally.[5] Further, nonverbal cues do not have universal meaning. They carry different meanings in different cultures.[6] Much of language then is nonverbal, and many thought processes contributing to language are nonverbal.

It follows that instruction and assessment that do not take these differences among students, any students and especially LEP students, into account are likely to be unreliable and invalid much of the time. Learning style and nonverbal language influence language and assessment responses.

4. The two language systems possessed by bilingual students limit the value of assessment methods used currently. In her article in the ERIC/CUE Digest, Carol Ascher[7] concludes that individuals who are bilingual have two distinct but overlapping language systems that they rely on in different ways depending upon the situations in which they find themselves. Because of this phe-

nomenon, she is particularly concerned that "diagnostic protocols" for bilingual students include information beyond standardized test scores and that assessments more directly aligned with curriculum be developed. Ascher's points are important. If bilingual students change language systems with the situations and stresses that confront them, we can never be sure which language system has interpreted (or misinterpreted) the multiple choice test item and produced the response which we are scoring. Assessments that enable us to know what language system is at work are needed.

Performance assessments, particularly portfolio assessments, have much to offer in assessment of LEP students. Potentially, they can contribute much more knowledge than we now are obtaining about what these students really know and are able to do. They offer potentially greater validity and reliability than present testing technologies. Many portfolio entries can be done in the native language, thereby reinforcing bilinguality and accommodating language system shifts. Since portfolio entries need not be restricted to print, these assessments can accommodate differences in learning styles and nonverbal communication. However, none of these possibilities can become realities unless those desiring to use portfolio assessments understand (a) what student portfolios are, (b) how they can be used, and (c) how to design them.

Student Portfolios: What are they and how can they be used?

What Is a Portfolio?

Current work in developing performance assessments focuses on three assessment protocols or types: portfolios, performance tasks and exhibitions. Figure 1 provides definitions of each assessment type and a few key issues in their development and use. Careful study of the definitions in Figure 1 should enable the reader to see where and how these three types of performance assessment might overlap.

A portfolio might contain a number of performance tasks or assessments of those tasks. In many cases, performance tasks require construction, creation, description (written or oral), or other formats for task completion that lend themselves to portfolio inclusion. Often, performance tasks are quite structured in time and space. In a recent joint proposal with Educational Testing Service to develop performance assessments, we defined a performance task as any reality-based task which would require an hour and a half or less to complete.

An exhibition could include presentation of a portfolio of work, although that need not be the case. Or, a portfolio might contain assessments and photographs or other documentation of an exhibition. Obviously, an exhibition is a display of what has been produced over time. The emphasis is on display or presentation.

The reader should also be aware that performance assessment can take forms other than portfolios, performance tasks, or exhibitions. Instrumentation used in the performance evaluations of teachers and administrators historically has included observation records, interview protocols and self-reports. Similar forms of assessment can be used in evaluation of student learning and could be included in portfolios. In addition, there is considerable effort at this time to use computer-simulated tasks as substitutes for "real" tasks which often require substantial equipment and/or materials for each student being assessed. It appears that the same tasks transferred to computerized formats are more efficient and cost-effective while losing little or nothing in their validity, reliability, credibility, or effectiveness.

How Can Portfolios Be Used?

While the definition of a portfolio provided in Figure 1 is French's definition, it is very close to the definitions of others active in portfolio development and utilization. Dennie Palmer Wolf,[8] a research associate with Harvard's Project Zero, defines a portfolio as "a chronologically sequenced collection of work that records the evolution of artistic thinking." Paulson, Paulson and Meyer[9] define it as "a purposeful collection of student work that exhibits the student's efforts, progress, and achievements in one or more areas." These definitions immediately suggest certain attributes of a portfolio that may be helpful in assessing student progress. Note that they emphasize a collection of work(s), chronological organization and purposeful construction (i.e., construction with a goal or purpose).

Additional concepts important to the formulation of portfolio structures and uses are offered by several researchers and developers. Howard Gardner,[10] director of Project Zero, suggests that portfolios can best be used to assess a student's ability to produce, perceive, and reflect. Wolf[11] describes portfolios as contributing "biographies of work" (e.g., a biography of the development of a musical performance), ranges of works (e.g., a collection of diverse pieces) and reflections (student analyses of what they have produced). Resnick[12]

Figure 1
Performance Assessments

I. Portfolio

A purposeful, Chronological collection of student work, designed to reflect student development in one or more areas over time and student outcomes at one or more designated points in time.

Key Issues:

- assessment targets/exemplars/performance standards
- guidelines for inclusions
- scoring/rating procedures
- training of faculty

II. Performance Tasks

A reality-based task which can be completed within the confines of a single day or less.

Key Issues:

- realism of the task
- scoring/rating procedures
- performance standards

III. Exhibition

The presentation of a body of work which has taken place over several weeks, months, or years.

Key Issues:

- realism of the task(s)
- scoring/rating procedures
- performance standards
- guidelines for development

NOTE: Defenses or reflections by the student(s) are often used in combination with all three of the above listed assessments.

compares portfolio development and assessment to scouting in that students use the same process as "accumulating badges over a period of years," i.e., they will complete tasks and submit projects that they wish to use to demonstrate competence against published criteria. Many teachers and others involved in examination of student portfolios mention the insights produced by portfolio entries about student learning, both what is learned and how it is learned. These comments often focus on student demonstration of communication skills, psychomotor skills, artistic skills and thinking skills as well as knowledge acquisition. Learning process dimensions discussed include critical thinking, socialization, perseverance, self-criticism, on-time task completion, problem-solving strategies, pursuit of quality or high standards and student ability to pose and address meaningful questions.

Clearly, there exists in the current literature the notion that a portfolio has the capacity to, and should, produce a portrait of both learning outcomes and learning processes, a portrait that enables the viewer (assessor) to see what the producer (student) is capable of doing and how he/she thinks, works, develops. Assessment potential is both formative and summative.

Some of the criticism of current and historic student assessment practices is also useful in determining what the role of portfolios in future assessment models might be. In his recent Phi Delta Kappan article, Stiggins[13] bemoans the state of "assessment illiteracy" among American educators. He defines "assessment literates" as persons who can recognize that assessment targets are unclear, that assessment methods are missing their targets, that samples of performance are inadequate, that there are specific extraneous factors creeping into assessment data and that assessment results are unclear. He calls for programs to train educators at all levels to be "assessment literates," thereby enabling them to create new forms of student assessment which are more valid, reliable, and appropriate.

Wiggins,[14] like Stiggins, expresses a concern for the identification of clear assessment targets. However, he refers to those needed targets as standards which he defines as "educative, specific examples of excellence on tasks we value." Current student assessments lack these "concrete" benchmarks (or exemplars) for judging student work at essential tasks, Wiggins posits.

In this context, the measurement of student progress toward the exemplar or standard requires a series of successive approximations. In other words, what's missing in both large scale and local student assessments are clear, specifications of exit level results against which student work is continuously compared.

The assessment model being promoted is criterion-referenced rather than normative, longitudinal rather than periodic, and output rather than input driven. This reliance on output, particularly exit level outcomes, implies that student work might take several varied forms to which a common set of standards (criteria) must be applied. Common standards can be applied only to completed products, tasks, or performances, Wiggins argues.

There are sufficient implications for assessment by portfolio in the Wiggins and Stiggins articles to round out our conceptualization of the role of this device. Notice that a portfolio has the potential to display various stages of student progress toward a clearly defined standard/exemplar/assessment target if one is defined. It offers a longitudinal assessment method that can be closely matched to the assessment target. It offers a means of collecting multiple samples of diverse kinds of student work and results (products) which are concrete and usable in a variety of ways.

The focus of this discussion of portfolios has been their role in performance assessment, and that will continue to be the focus in the remainder of this paper. However, it should be noted that portfolios are often used as instructional devices. In fact, one of the current problems in portfolio development and utilization is the tension between instruction and assessment that many classroom teachers seem to feel.

Although assessment should be aligned with instruction, and assessment results should be used to direct subsequent instruction, these two processes have different parameters and requirements. Good teachers have long used monitoring (informal assessment) of student activities to make immediate adjustments in student tasks and in their own instructional practices. Further, they often feel obligated to give immediate assistance to students struggling with a task. Neither practice is appropriate to summative assessment in which validity and reliability of measurement must be maintained. When the purpose of the task or exercise is to determine a student's accomplishment of a prescribed standard or to determine progress toward that standard, students must be allowed to complete and submit products or productions for scoring without additional assistance. After assessment is completed, reteaching, additional review or additional practice can take place. It appears that portfolios tend to blur the critical lines between instruction and assessment and between formative and summative assessment even more than present testing procedures for many educators.

What Should A Portfolio Contain?

There is no simple answer to this question. Obviously, the type of portfolio, its storage and retrieval system, the subject areas, skills, and processes involved in the assessment and the characteristics of the student(s) have to be considered in determining type and number of portfolio entries. Currently, various portfolios include written materials (essays, stories, themes, compositions, research papers, etc.), anecdotal information (logs, journals), work samples (selected seatwork, homework), projects/products (things created by the student or representations of them), tests/test scores, teacher comments/analyses, self-analyses, audiotapes, videotapes, photographs, drawings, paintings, observational records, and checklists. Notice that some of these items are the products of student activity, and some are assessments of student activities. However, the elements of the definition of a portfolio should be kept in mind. It is not a random collection of whatever is available, but a chronological collection of artifacts carefully chosen to represent the student's achievement of specified objectives and/or progress toward them. The outcomes being measured may be acquisition of knowledge, cognitive, psychomotor or social skills or attitudes and dispositions.

While the system for storing and retrieving information in a portfolio plays a significant role in determining types of entries, the limitations of space, time, and format are swiftly being erased by the technology now available. Linda Vista Elementary School in San Diego, California, has been experimenting for two years with a computerized portfolio that allows for computer storage and retrieval of multiple types of information including print, videotape, voice prints, and photographs. It is also interesting to note that Linda Vista Elementary School, an RJR Nabisco Foundation Next Century School, has more than 60 percent LEP students representing six native languages: Hispanic, Vietnamese, Cambodian, English, Laotian and Hmong. Students in the school are not grouped by age or grade level but by English proficiency, and aspects of the curriculum are taught in each native language. Obviously, the electronic portfolio is perceived as a means of accommodating a range of learners and languages and gathering data for assessment which transcends the boundaries of current standardized tests.[15]

Uses of Portfolios in Student, Teacher and Program Assessment

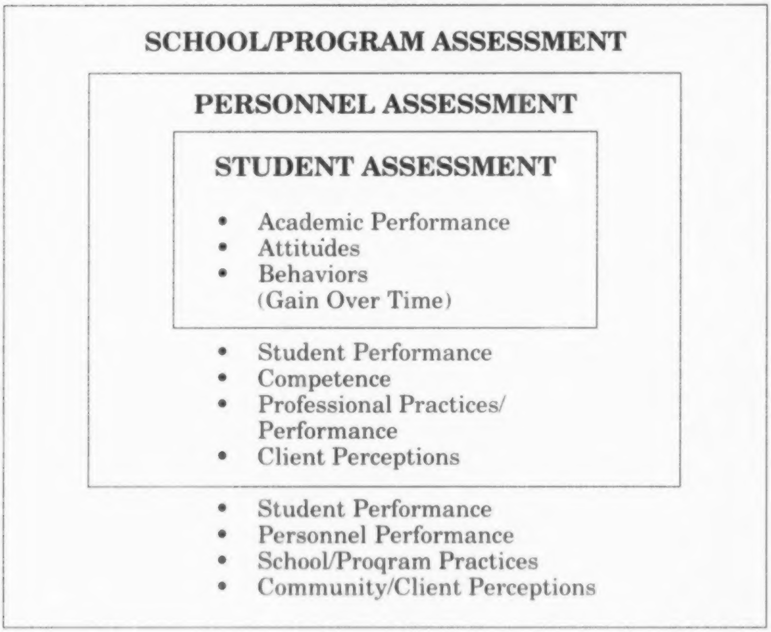
Figure 2 presents a detailed summary of a nested assessment model. At the center is student assessment; at the second level is personnel assessment, and at the third broadest level is school or program assessment. Each assessment level builds upon those encompassed within it. Clearly, student assessment is or should be at the core of the model, and portfolios can play a major role in all three aspects of the assessment.

As one surveys the model, it should be remembered that it assumes that the only appropriate focus for any of the three levels/

types of assessment is gain or change over time. That means that the first step at any level is to produce baseline data at some initiatory point (beginning of year, beginning of school, initiation of a program) against which performance at other points in time can be measured. Consideration of that proposition leads quickly to an understanding that obtaining baseline data about student performance is critical to the whole assessment model. Portfolios offer one means of capturing baseline data and adding data over time which can clearly show gain or change.

Figure 2

ASSESSMENT IN PERFORMANCE MANAGEMENT



SCHOOL/PROGRAM ASSESSMENT

PERSONNEL ASSESSMENT

STUDENT ASSESSMENT

- Academic Performance
- Attitudes
- Behaviors
  (Gain Over Time)

- Student Performance
- Competence
- Professional Practices/ Performance
- Client Perceptions

- Student Performance
- Personnel Performance
- School/Proqram Practices
- Community/Client Perceptions

Student Assessment

As shown in Figure 2, student outcomes should be defined more broadly than scores produced on achievement tests. In many cases, student attitudes and self-management behaviors must be changed before academic performance can improve. In some cases (e.g., severely handicapped students, preschool age children), attitudes and behaviors are the teaching-learning focus rather than academic content. Therefore, measures of student outcomes should encompass academic outcomes, attitudinal changes and something which might be labeled intellectual growth.

In discussing the uses of portfolios in student assessment, Howard Gardner suggests that they lend themselves to assessment of products, perceptions and reflections. In that framework, measurement of academic outcomes might be seen as assessment of what the student is able to produce.

When measuring changes in attitudes, we are usually measuring changes in students' perceptions and feelings, changes which may be very important to what and how they produce. Gardner, Wolf, and others involved in portfolio assessment have found that assessments of this type offer rich insights into student perceptions at various points in time, if portfolio designs require the inclusion of materials that can be reviewed for this dimension of performance.

Obviously, I have defined academic outcomes and intellectual growth differently. It may not be a very valid separation, but the term "intellectual growth" is used here to try to identify the potential of what Gardner has called student "reflections." What and how students thinks about their own work, progress, growth, development is the focus. One might talk about this area as thinking skills, but thinking skills are essential parts of the other two areas identified for assessment as well. Foremost, the separation of this area from the others is meant to suggest that student reflections will not be forthcoming unless they are designed into the assessment methodology.

Personnel Assessment

Personnel (teacher, administrator) assessment should focus on student outcomes, but if one does not know what inputs produced the outcomes, there is little chance of improving outcomes, especially school-wide outcomes. Therefore, what the professional educator knows and is able to do (competence), his/her application of effective teaching or administrative practices (i.e., practices proven to produce higher outcomes), and the satisfaction of those for whom he/she is responsible (an important ingredient in classroom and school climate) are also important focuses of assessment.

Instructional practices are evidenced in portfolio collections. Commonality of student approaches to problem solving result from teaching not inspiration. Systematic errors in written work across a class of students or a school reflect instruction. Portfolio entries made by teachers and comments by teachers on student entries provide insight into instructional values. The types of tasks and projects included in portfolios speak of instructional methodologies as well as curriculum content. Hiebert and Calfee[16] suggest that student portfolios provide links between instruction at several grade levels. If the linkage is there, analysis of portfolios at several grade levels should demonstrate it.

Program Assessment

When one changes the assessment lens to focus on the quality and success of a program or school, portfolios also contribute in a variety of ways. If we think of programs in terms of inputs, processes, and outcomes, it becomes easier to see where and how these contributions are made.

Program inputs are usually defined as goals and objectives, characteristics of the target population, available resources (human and fiscal), facilities, organizational structure, and other such variables.

As we consider inputs, Grant Wiggins[14] makes an interesting statement about schools and standards:

> A school has standards when it has high and consistent expectations of all learners in all courses. High standards, whether in people or institutions, are revealed through reliability, integrity, self-discipline, passion and craftsmanship.
>
> Alas, it is thus not too strong to say that many schools exhibit no standards.

If objectives/specifications/standards/exemplars/desired outcomes are clearly defined for student portfolios, much can be learned about the program expectations and goals. Are the standards/exemplars high? Are they short-term? Longitudinal? Have exit standards been established? The absence of these elements also tells us much.

Study of portfolio specifications and guidelines also tells us a great deal about target audiences. Wiggins[14] argues,

> If we are to obtain better quality from schools, we are going to have to challenge the current low expectations for all students in a course, age-cohort, and entire school population.

Are the portfolio standards/exemplars for all students? Are there differentiated standards? How much variance in performance will be allowed? For whom? Initial portfolio plans (they may change over time) also contribute information about school and program organization. Who can enter materials or comments? Who contributes to assessment? What kinds of entries can be included? Responses to these issues provide insight into faculty and subject matter organization and student involvement.

When assessing the results and impact of programs, it is at the input stage that extensive data regarding pre-conditions should be collected. Since most portfolio designs call for the collection of student work samples at the beginning of the year or portfolio initiation, analysis of the quality of these samples across portfolios offers some information about the state of curriculum, instruction, and learning prior to program initiation as well as a baseline against which to measure the progress made by all students over time.

In program assessment, processes include elements such as curriculum, instructional practices, parent, and community involvement and professional development of educators.

Portfolio contents, when viewed collectively, give great insight into curriculum emphases. There is tangible evidence of subject matter knowledge learned and/or emphasis on communication skills or thinking skills or artistic skills or problem-solving or whatever other emphases have been consciously or unconsciously stressed. Where individual differences in student learning styles or interests or ability have been consciously addressed by program staff, a survey of student portfolios should confirm that. When conscious attempts have been made in a school or program to integrate disciplines and subject areas, portfolio entries can provide evidence of the results.

Lorrie Shepard[17] argues that better student assessments are needed because current tests narrow the content taught. In other words, curriculum tends to focus on what is tested. She also argues that the content of all assessments must be negotiated at some level or another. Therefore, what appears in an assessment represents some kind of consensus building process regarding curriculum. If portfolios are being used as student assessment devices, a survey of their contents should indicate whether curriculum content is narrowing or expanding. Further, portfolio specifications, guidelines, and contents should alert the program assessor to the levels and types of curriculum consensus that have been or are being built.

The contributions of student portfolios to personnel assessment have already been discussed. We can simply reinforce here the notion that at the program level the emphasis should be on instruction not individual instructors. To determine the quality of and emphases in instruction, we must look across portfolios not within a single portfolio.

Student portfolios may or may not provide information about school environment or parent involvement. It depends upon the types of information collected and placed in the portfolio. For example, several of the Next Century Schools projects are collecting periodic assessments of student self-concepts. However, these may not become part of a student's portfolio.

Student portfolio contents may not give much insight into the professional development of teachers and administrators, but the presence and design of the portfolios can. Hiebert and Calfee[16] conclude that "student portfolios provide vivid and engaging content for professional discussion and collegial sharing." The conclusion is supported by this author's experience. Successful portfolio assessment projects that were designed without this dialogue and sharing appear to be non-existent. Successful projects, in which professional dialogue among program/school staff about the quality and meaning of portfolio entries is lacking, also appear to be very infrequent, if not non-existent. The presence of student portfolios offers the program assessor several avenues for dialogue with administrators and teachers about the professional growth and development that is taking place.

Since the primary function of student portfolios is assessment, their presence and contents should provide the program assessor with direct information about the alignment of curriculum goals, instructional strategies, and assessment activities. Collecting information about these alignments has long been an issue and intent of curriculum evaluation.

Program outcomes are inclusive of student academic achievement, affective development, attitudes and behavior, teacher and administrator morale, and changed school/program organization. When assessing program outcomes, student portfolios should contribute greatly. If clear performance standards/assessment targets have been created, individual and collective student achievement against those standards can be readily measured. Progress of students of different types and levels should be easily identifiable.

The "biography of a work" which Wolf describes as a product of portfolio development can, in program assessment, be translated to a biography of students' works in which one can read a number of outcomes. Perhaps one of the most important outcomes at the program level will be the consistency of performance across students. If the challenge to low expectations for all students in an age-cohort, class, or program for which Wiggins argues has been mounted, differences in student performance outcomes should be minimal; i.e., they should be within narrow, tolerable limits.

Student affective and attitudinal changes as well as academic progress can be assessed to some degree in the construction, characteristics, and quality of the work produced over time. If student perceptions and reflections as well as performance are valued and developed, evidence should exist in portfolio contents. "By looking across portfolios, we begin to see where people excel or flounder," as Wolf[8] contends.

If student portfolios are multi-year endeavors, their contents, specifications, and guidelines are bound to change as the professional staff involved with them change and grow and as the school or program organization changes. Analyses of these changes offer insight into organizational and professional outcomes as well as student outcomes.

It may appear that more attention has been given to the uses of student portfolios in program assessment than to their uses in student assessment. Certainly, more space has been given. However, the approach here was a conscious, purposeful one. Many of the questions about portfolio application to student assessment should be answered in the next section of the paper devoted to portfolio design issues. In addition, my review of several reports of LEP program evaluations indicated that these program assessments were superficial at best. An attempt has been made in these last few paragraphs to suggest ways of collecting and analyzing data from student portfolios which can be of much use in determining the quality and success of an LEP program.

Portfolio Design Issues

Past And Present Problems

The development of portfolios of student work and learning products will be of little value to formal student assessment unless portfolio structure, contents, and evaluations of contents are carefully designed before portfolio development is undertaken. The problems of the past must be resolved.

Historically, attempts to use portfolios in assessment have met with six problems. Expectations (objectives) of those conducting instruction and assessment have been unclear to both evaluatees (in this case, students) and instructors/evaluators. Guidelines for number and type of inclusions have been nebulous or non-existent; thereby, reinforcing evaluatees' beliefs that "if some inclusions are good, more are better." The results are sizeable, uneven, unequal, and sometimes unrelated stacks of materials and products constituting evidential bases for assessment decisions. Procedures for scoring or rating portfolio entries, combining assessment results, and clearly communicating student outcomes to students and parents have not been clearly thought out and communicated to those who need to know. A clear decision about the measurement construct to be used in analysis of portfolio entries and use of those results has been lacking; i.e., "Are portfolio entries to be used in a criterion-referenced evaluation context (student development over time) or a normative context (comparison of accomplishment among students)?" Entry and analysis procedures have been unclear; i.e., questions such as the following have not been thoroughly discussed and resolved in advance of implementation of the portfolio process:

- Who (students, teachers, others) can enter materials?
- Who (students, teachers, others) participates in assessment? How often? Under what conditions? What standards will be employed?
- For what period of time will portfolio entries be kept? For how long are they valid indicators of progress or accomplishment?
- Who has access to the portfolio and the evaluation results?
- What procedures will be used to delete entries from the portfolio, when and if necessary?

Persons given the task of evaluating portfolio entries have been given little or no training in how to evaluate them and few standards against which to measure progress. The results are high inference and subjectivity.

Portfolio Design Questions

The questions below can form the skeleton of portfolio design. Designers may wish to add others that address uniquenesses in their students or settings.

1. What instructional goals, objectives, and outcomes do we want to measure?
2. Which ones (goals/objectives/outcomes) are not now being assessed adequately by other means?

   NOTE: Don't reinvent the wheel. If current assessment methods are adequate, why switch?

3. Will portfolio entries and their analysis be used to assess individual student progress over time or to compare student accomplishment taking into account individual differences?

NOTE: Will the portfolio be used for criterion-referenced assessment or normative assessment? The answer will dictate much about types of entries and procedures for entry.

4. What evidence of progress and/or accomplishment will be required? What evidence of progress/accomplishment will be allowed?

   NOTE: The first question addresses the need for a consistent base of information from student to student. The second addresses issues of individual differences such as creativity, best effort, learning styles.

5. Who will select entries? Why?

   NOTE: In some plans, teachers select all entries. In others, students build their portfolios within specific guidelines. Several researchers and developers recommend that both parties be contributors. What about administrators? Parents? Obviously, age of students, content area and other factors need consideration.

6. What types of evidence can/will be accommodated in the portfolio? Why?

   NOTE: This question was addressed in an earlier section of the paper where it was stated that the type of portfolio, its storage and retrieval system, the area(s) of content involved in the assessment and the characteristics of the students have to be considered.

7. How will portfolio contents be rated/scored/judged? Used in student valuation? Program evaluation? Instructional improvement?

   NOTE: These questions require resolution of both measurement and evaluation issues.

8. Who (students, teachers, others) will contribute to the assessment?

9. How will assessors be trained? What controls will be used to assure some degree of validity and reliability in assessment results?

10. How will results of portfolio assessments be communicated to students? To parents? To the school district?

11. How and when can/will portfolio entries be deleted?

12. What can/will we learn about the success of our program/project from the analysis of student portfolios?

Issues For Discussion

Underlying the twelve questions above are a number of philosophical and measurement issues that need to be discussed and some agreement reached by the professionals in a program, school, or school district before portfolio utilization is undertaken. Perhaps the discussion is best facilitated by development of propositional statements such as those below which are offered for debate. They are a compilation of many of the premises found in the current literature on student portfolios.

1. Portfolios can best be used to assess a student's ability to produce, perceive, and reflect.

   NOTE: This statement is attributable to Howard Gardner, Harvard University (see references).

2. Portfolio entries should be selected by both students and teachers by mutual agreement. Both parties have a stake in the teaching/learning process.

3. In program assessment, portfolios provide insight into process as well as products and outcomes.

4. Portfolios are best used to assess student development over time rather than to assess comparative accomplishments of students.

5. In the arts and humanities, the versatility of the student should be assessed.

6. Portfolios do little to accommodate learning styles unless students are encouraged to produce and submit diverse types of materials and products.

7. Portfolio development and cooperative learning activities go hand-in-hand. (The two can be easily related.)

8. In areas such as writing, evidences of the whole process are more useful than the final product(s) alone.

9. If student reflection is desired, both self-critiques and teacher critiques of entries are required (so that teachers and students can compare them).

10. Evaluation of portfolio contents requires at least two levels of organization: categorical organization of raw data/evidence and summaries or syntheses of available data.

11. If portfolios are to be used in assigning grades, scale descriptions for the requirements for A, B, C, D etc. must be developed.

At least two additional propositions for debate among those contemplating portfolios for LEP students should be added to the list:

12. Some, but not all, written and oral portfolio entries should be in the student's native language. The choice of which entries will be in English and which in the native language should be the student's.

13. Raters/scorers of portfolio entries by LEP students must include at least one person proficient in the student's native language.

Scoring/Rating Portfolios of LEP Students

Rating or scoring portfolio entries may be as simple as scoring a set of responses to a mathematics quiz in which problems have right or wrong answers; i.e., some entries may be sorted on the basis of right or wrong, accurate or inaccurate. However, that often is not the case. Many entries require the exercise of professional judgment. For example, musical compositions, photographic essays, various pieces of writing, and videotaped performances require more of the rater and rating system than has been typical in many testing programs. The issues are compounded when limited language proficiency and/or the use of multiple languages are added to the situation. At least six elements are needed to properly conduct the scoring/rating process.

As indicated in earlier comments, standards for performance must be predetermined when rating portfolio components. What constitutes an outstanding performance? An acceptable performance? An "A"? The standards should be, as Wiggins[14] suggests, exit standards; i.e., they must be standards that describe acceptable performance at the end of the educational process. In some cases, those standards will be end-of-the-year standards. In others, they may be school exit standards. Acceptable progress toward those exit standards should be judged in terms of movement along a continuum from each student's entry point to the exit standard. Portfolio entries of all students should be judged against prescribed standards, not against each other. For LEP students, the performance standards should address language standards as well as other elements.

Both students and teachers need exemplars of performance at the prescribed standard. What does an outstanding musical composition look and sound like? An acceptable short story? An award winning photographic essay? If native language or mode of thought is to be used, exemplars in the language need to be provided.

Many products and performances probably will be rated on a scale of some sort. This tends to be true even when numbers of points are awarded for the presence of certain features in the product/performance. Usually, the points are totaled and applied to some pre-determined scale. (Our typical grade structures operate like this.) Scales used in rating entries should be behaviorally anchored scales; i.e., each point on the scale should be described in terms of the behaviors required to achieve that level. What elements of performance must be present to achieve a "5" (on a five-point scale)? What elements can be absent and still allow the producer to obtain a "3"? If exemplars of exit level performance have been provided (e.g., writing performance at the end of the high school years), what elements must be present to obtain an outstanding ("5") rating at the end of the middle school years?

A fourth element necessary to rating and scoring is the use of multiple raters/evaluators. Olympic competitions rely on multiple judges. If portfolios are to be a serious part of student assessment, an approach not unlike that used to score Advanced Placement Examinations should be used. A team of raters (at least two) will add validity and reliability to the assessment score. Further, the use of multiple raters is essential in assessing portfolios of LEP students. If native language is allowed, one or more members of the rating team will need to be proficient in the native language. If entries make use of only the English language, there is still need for at least one rater to be proficient in the native language. He or she will be the person more likely to identify the characteristics of the product or presentation directly attributable to language and bring these to the attention of colleagues.

Although it may not always be essential, this writer recommends the use of consensus processes among raters. Rather than supplying two or three independent ratings/scores which are then averaged, each rater generates independent ratings, then meets with colleagues. Ratings and rationales are shared, and the group arrives at a consensus rating and a consensus rationale for that rating. While this approach requires additional time, it strengthens validity and reliability of the final scores, contributes to the comfort and "assessment literacy" of raters, and provides staff development both in assessment and instruction. Rater teams always seem to talk about what can be done to improve student performance.

If the reader has followed closely the five rating/scoring process elements described thus far, he/she can predict the sixth. Raters/scorers must be trained. They must be trained in how to apply the standards, exemplars and rating scales to the student products. If consensus is to be used, they must be trained to use the consensus process.

A Final Comment

New student assessment technologies, including portfolios, can provide new and often better information about student performance and development and about program performance than has previously been available. There appears to be great potential in the use of portfolios with LEP students. However, the value is yet to be determined. Experimentation, perhaps as much as ten years of it, will be needed. Thankfully, that experimentation is underway.

Notes

[1] National Governor's Association. (1991). From rhetoric to action. Washington, D.C.: Author.

[2] National Business Roundtable. (1990). Essential components of a successful education system. Washington, DC: Author.

[3] National Center On Education And The Economy. (1990). Seven components of the restructuring process. Washington, DC: Author.

[4] French, R. L. Teaching strategies and learning processes. Tennessee Education.

[5] Mehrabian, A. (1971). Silent messages. Wadsworth Publishing Company.

[6] Hall, E. T. (1969). The hidden dimension. Anchor Books. Additional information can be found in Hall's works, The Silent Language and Beyond Culture.

[7] Ascher, C. (1990). Assessing bilingual students for placement and instruction. ERIC/CUE Digest, (No. 65). New York: Clearinghouse On Urban Education.

[8] Wolf, D. P. (1988, December, January). Opening up assessment. Educational Leadership, 45(4), 24-29.

[9] Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991, February). What makes a portfolio a portfolio? Educational Leadership, 48(5), 60-63.

[10] Brandt, R. (1990, March). On assessment in the arts: A conversation with Howard Gardner. Educational Leadership, 47(6), 24-29.

[11] Wolf, D. P. (1989, April). Portfolio assessment: Sampling student work. Educational Leadership, 46(7), 34-40.

[12] Resnick, L., & Associates. (1991). The New Standards Project: An Overview. University of Pittsburgh, Learning Research and Development Center.

[13] Stiggins, R. J. (1991, March). Assessment literacy. Phi Delta Kappan, 72(7), 534-539.

[14] Wiggins, G. (1991, February). Standards, not standardization: evoking quality student work. Educational Leadership, 48(5), 18-25.

[15] For more information about Linda Vista Elementary School and its electronic portfolio, contact Dr. Adele Nadeau, Principal, 2772 Ulric Street, San Diego, California 92111.

[16] Hiebert, E. H., & Calfee, R. C. (1989, April). Advancing academic literacy through teachers' assessments. Educational Leadership, 46(7), 50-54.

[17] Shepard, L. (1989, April). Why we need better assessments. Educational Leadership, 46(7), 4-9.

Response to Russell French's Presentation

Alice J. Kawakami
Pacific Regional Educational Laboratory, Hawaii

I would like to thank OBEMLA and OERI for inviting me here as a discussant. Although most of my experience in education in Hawaii has not been with bilingual education, Hawaii's public school population is multi-ethnic and students come from various language and dialect backgrounds. I have been involved in developing a language arts program which has been attempting to develop portfolios as a means of assessing elementary students' learning. With this background, I find Dr. Russell French's paper insightful and important in focusing on some of the key elements of authentic assessment within a classroom setting.

As I consider the theme of this conference, "Achievement as a Child's Universal Language," it seems that one way for our students to participate in that universal language and to be able to speak of their achievements is to provide them with opportunities to give voice to their successes through the use of portfolio assessment. As Dr. Russell French points out, current standardized tests do not adequately "showcase" the learning of our students. I will speak today on my experiences with the Kamehameha Elementary Education Program (KEEP) in Hawaii, during the development of portfolios. KEEP is a language arts program for elementary students. It was developed to assist Native Hawaiian students in the public schools. In its early years, the program was grounded in culturally congruent interaction styles, classroom organization, comprehension focused direct instruction, and a mastery learning system to track student progress. Recently, the curriculum was expanded to maintain the cultural component and include the development of writing and, most importantly, to focus on the development of students' ownership of their learning. This required a paradigm shift in curriculum and assessment as well as on the part of the teachers. The relevance of that work to our topic today is the process of moving from assessments which were standardized to authentic assessment. The focus of my comments will be the support needed for teachers undertaking this change in teaching and learning and the documentation of more functional, authentic learning.

With this change, instruction and assessment took on a different, more responsive face. The transition from a reliance on test results to portfolio assessment is still in progress. Many of the issues raised in Dr. French's paper were addressed in the process of program development at Kamehameha. I would like to describe some of the major changes and accompanying support that was needed. Moving to portfolio assessment was not easy because it called for a reexamination of our basic beliefs about the role of teachers and students and the criteria by which we measure success.

In retrospect, we found the framework of a paradigm shift helpful in understanding the changes we have made. The framework enabled us to look at some of the assumptions underlying the original program (based on a transmission model of instruction) and the current whole literacy program (based on a constructivist model of instruction). The following table outlines four areas of change. The content and process of instruction refer to the actual classroom routines. Assessment and evaluation refer to the monitoring system that directs the focus of instruction. Accompanying these routines in the classroom are perceptions about the roles of the teacher and the student.

A Paradigm Shift in Teaching and Learning

	Transmission Model	Constructivist Model
Instructional Content/Process	predetermined classroom content	constantly constructed content within class
Assessment	external criteria	internally developed criteria
Role of teacher	giver of knowledge	facilitator of learning
Role of student	receiver of knowledge	coordinator of own learning

Under a transmission model, content and instructional strategies are fixed by curriculum guides and published materials. In a constructivist model, content is based on curriculum areas but negotiated with student input on topics of interest. Assessment in the transmission model is usually dependent upon externally set criteria such as skill-based mastery tests. Assessment in the constructivist model is based on goals set collaboratively by the teacher and student as the criteria for success. This criteria arises from the context of the classroom and is tied to benchmarks for student progress. It provides feedback for learning, and is useful to students, teachers, and parents. This is the critical function that portfolio assessment provides in the KEEP whole literacy classroom. It is responsive to the classroom learning environment and should be authentic and meaningful. These models of teaching and learning are supported by the roles of teacher and student as indicated in the table.

In order for teachers to support student learning through the use of portfolio assessment, a shift must occur in their perception of their role as teacher. This shift may be difficult for many teachers because it is predicated upon change in the assumptions about teaching and learning. Support for making this shift is critical because it calls for major philosophical change. All change is difficult to bring about and this is not an exception. Our experiences at KEEP taught us that these changes cannot be mandated but need to be developed. The development is dependent upon the extent to which staff development activities can be grounded in the same constructivist model.

The following change matrix is helpful in understanding the change process and the kinds of support needed by teachers. These changes must be viewed as part of the process of developing portfolio assessment. With tests as assessment, the guidelines for administering the tests were made explicit in the testing procedures. With portfolio assessment, guidelines for developing a portfolios are grounded in the teacher's attitudes, beliefs, and values which require collaboration with students. The curriculum, classroom interactions, teacher development, and student development are an integral part of the implementation of portfolios.

In moving toward a classroom environment where the teacher acts as a facilitator of student learning and the students take responsibility for their progress, teachers move from a relative position of isolation and reliance on curriculum materials and guides to a context of professionalism and collegiality. The following change matrix is useful in understanding the steps involved in moving from one point to another within the framework. Through our often frustrated efforts, we found that institutional support for teacher development along these lines is critical.

Change Matrix

Content	**Context**		
	Isolation	Congenial group	Professional colleagues
Materials	A		
Behavior/ Strategies		B	
Attitudes, Belief, Values			C

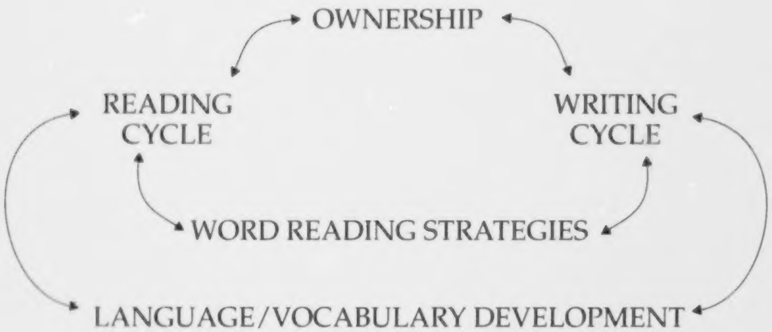
We can view the process of change on three points in this matrix. Initially, at point A, teachers may find themselves operating in the isolation of their own classrooms. Their primary guides for professional development are the instructional materials used in their classroom. This is ideal for teachers operating under a transmission model, with fixed scope and sequence charts, lesson plans and assessment instruments developed by publishing companies far from the contexts of their classroom. As a teacher begins to seek the input from colleagues, point B may describe the interactions. Here, teachers meet in groups and discuss students' behaviors and discuss strategies they have tried. This stage lends itself to more dynamic views of teaching, and more experimentation based on classroom conditions. A teacher who is operating in a constructivist model would be placed at point C, Here, attitudes and beliefs about teaching and learning, strategies for teaching and instructional materials are discussed with colleagues. Teachers at this stage reflect upon their teaching, seek feedback from others, and take responsibility for their own professional development. This stage is analogous to students who have assumed responsibility for their own learning.

When we first began our training, we thought that by giving teachers whole language instructional materials, we would instantly make them constructivist teachers. We created opportunities for groups of teachers to meet and discuss materials and that is exactly what we obtained, groups of teachers looking at big books. Our next step was to provide them with training in teaching strategies. We provided workshops on writing process approach and shared readings. Still, change was negligible for the majority of teachers. We finally realized that to effect long-term, deep change in teachers, we needed to address their attitudes and beliefs. We took the approach of focusing on those underlying assumptions in day-long retreats. We posed questions for discussion about the learning process that compelled teachers to examine their own learning process, their beliefs about the best learning environment for their students, and the most appropriate role for them to take vis a vis their students. At this point, we began to see progress in developing reflective, responsive teachers who began gaining confidence in their abilities to function as facilitators of student learning. Until that moment, the need for portfolio assessment was not realized.

In our discussions of shifting the focus of student learning goals, we had to develop an innovative means of conveying the objectives of the language arts program in a way that was not tied to a format with the limitations inherent in a scope and sequence chart. We decided to begin by presenting teachers with a diagram of aspects of literacy to be developed within the program. There are six aspects of literacy in the KEEP curriculum. The following diagram shows the presentation format that we used.

KAMEHAMEHA ELEMENTARY EDUCATION PROGRAM

SIX ASPECTS OF LITERACY



VOLUNTARY LITERACY
(Reading/writing out of school)

The first and most important aspect of literacy is the development of students' ownership of their own literacy learning. This translates into activities which operate within meaningful contexts. The purpose of reading and writing activities in school is to communicate ideas that are relevant to the students rather than to complete a number of worksheets or to move through a specified number of pages in a practice book. This provides authentic literacy activities. If the content of the communication is meaningful to the students, instruction in word reading strategies and language and vocabulary development is purposeful as a critical part of communicating. If reading and writing in school are meaningful, the application of these abilities should flow beyond the realm of classroom work and into the realm of voluntary reading and writing. This voluntary literacy, outside of the school, is the application of school learning and a demonstration of students' ownership of learning. This concept of the curriculum as the basis for real application of school learning is the perfect context for the development and use of portfolios. It is also an absurd context for assessing learning through standardized tests.

Teachers who are committed to developing students' ownership of functional meaningful literacy recognize the artificiality of testing developed under a transmission model. However, until teachers are committed to this concept of literacy development and have recognized the attitudes and beliefs that underlie their own teaching, portfolios may be just another requirement of the curriculum. When teachers sit with their students and collaboratively document the kind of literacy learning that has taken place, portfolios take on the role of measuring the complex, functional learning that is needed for success in society. Evidence to document progress is far different from information on skill-based mastery tests. More appropriate documentation would be described in Dr. Russell French's categories of performance tasks, exhibitions, and portfolios. In KEEP classes, projects, reading logs, samples of a student's best writing, and correspondence between students, teachers, and parents qualify as legitimate assessment information. Benchmarks have been developed as ties to state performance expectations.

If bilingual students and students from multicultural homes are to meet success in our schools, assessment must be designed to allow for authentic learning to be a legitimate part of the definition of success. My comments reflect a perspective that portfolio assessment must be negotiated by teachers in classrooms and, for that to occur, teachers will need a lot of support in developing the attitudes, values, and beliefs that free them from the confines of standardized tests. All of the specific issues raised in Dr. Russell French's paper need to be addressed as portfolio assessment is implemented in schools. In addition to those concerns addressing authentic assessment of complex learning of students, there must be attention to the complex learning that teachers will experience as the face of assessment changes to reflect complex, functional learning. The problems that are identified in Dr. Russell French's paper:

> unclear expectations, nebulous or non-existent guidelines, unclear scoring procedures, lack of definition of the measurement constructs for portfolio entries, vague entry and analysis procedures, little or no training in implementing portfolios, and few standards were all a part of our process to shift to authentic assessment. Although all of these issues have not been completely resolved, teachers must be involved in the development of portfolios, because they will be at the delivery point of this new assessment format.

The technical questions raised in Dr. Russell French's presentation are vital to the development of authentic assessment. My remarks were intended to point to the equal importance of the teacher and the support that is needed to implement changes in assessment. Someday, our students will be able to use portfolio assessments to support their success in learning and allow us to recognize their real achievements.

Response to Russell French's Presentation

Daniel Koretz
RAND Corporation

A lot of the issues that have been raised so far pertain to portfolios in general not specifically. Their use is with children who are limited in their use of the English language. I'm going to skip over a lot of the issues that are more generic, but I do need to touch on a few to make my more specific comments clear. I should clarify at the onset that I'm speaking as a proponent of portfolio assessment. I've been involved in developing the portfolio assessment program for more than three years. I do think that, properly done, portfolios have a substantial potential for improving instruction, but I think that they're very difficult to do and, if we don't go into our efforts to use them with eyes open, we stand to lose more than we gain. Moreover, I think that it's very doubtful that performance assessments will provide the kind of opportunity that was alluded to twice in the past two talks, that is revealing abilities of LEP children that have not been revealed by traditional tests. I think they may represent a real opportunity to LEP children, just as they represent a real opportunity for any children, and that they may help steer teachers, as Dr. Alice Kawakami suggested, toward more interesting, engaging, and demanding course work. This pertains to all students, regardless of their native language. I'll come back to why I'm a little more skeptical about their usefulness in revealing abilities of LEP children that have been hidden because of their difficulties with English.

A few generic comments. First, portfolio assessment and performance assessment, in general, really have two different goals. The portfolio's assessment goal is one of getting better assessments of what children can do. Better in the sense of tapping abilities or capabilities that traditional tests might not, and performance assessment is improving instruction. Those two goals are very different despite the fact that proponents of performance assessment often talk of them both at the same time and often assume either implicitly or explicitly that, if a task or an assessment is authentic, and I'll leave open what that means, it will improve instruction and provide good assessment. I think that's simply not true in many cases. It's very easy to come up with tasks that are engaging in the classroom and potentially very useful in the classroom but have no discernable measurement value. There are some cases where they do overlap, but I think the overlap can easily be overstated. I think the conflict between those two goals or the uneasy compromise between those two goals is particularly severe when the children are being tested at limited proficiency in English, and I will come back to why that is so.

As a second generic caution, portfolio assessment is extremely difficult. It is not hard to come up with a collection of work from an individual student that the teacher and the student may agree has been properly evaluated. It is extremely hard to get collections of work from large numbers of students that are rated in a way that is even halfway comparable from teacher to teacher, school to school, child to child, which is what large scale assessment programs have to do.

I've been involved in the Vermont performance assessment work since it started more than three years ago. Most of the participants remain very enthusiastic about portfolios, very optimistic that they will, in fact, help improve instruction, but the list of difficulties that the participants have faced in the past three years is very long. I will just list a few of them for you. One is that because portfolios can include, as Dr. Russell French mentioned, a wide and diverse array of materials, raters often find that they get work in portfolios that they can't rate. Once they have agreed on standards, on criteria for judging student work, lo and behold, children produce things that don't fit.

The converse of this is raters who report that they periodically, in fact, not too infrequently, come across good work that they recognize as hard work from a capable person that slips through the cracks, because it was not the kind of work that was in mind when people designed the criteria. Raters reported that it was extremely difficult to aggregate the ratings of individual tasks because a portfolio is a collection of tasks and products submitted to some summary judgment of individual work. A particularly severe problem is that the nature of classroom assignments was often too poorly documented for the raters to judge performance. For example, if a student in a mathematics portfolio does not show adequate explanation of why he or she solved the problem in a certain way or how he or she came up with the answer, is that because the student can't do it, didn't do it, or because the teacher's assignment didn't give the student any reason to do it. Well, often, you can't tell. Let me put it differently. It is very hard to tell, and it requires a lot of careful work to make sure that the relevant documentation is there.

Sometimes raters found that they lacked enough information about students to judge what they were doing. For example, again in mathematics, a given solution to a specific problem for some students might be a remarkable act of invention if that student has never confronted that particular kind of solution before. But, if in fact, another student has worked on that kind of problem at great length and happens to know that one of the ways to solve a problem of this type is to use the such and such method and just regurgitates it back, that is, in a sense, a much lower or at least different kind of performance. How do you know which is which? I mention these not to discourage you but to encourage you not to see portfolio assessment or performance assessment, in general, as some kind of panacea. It's damned hard work, and it often doesn't work.

Now, one of the consequences of these problems is that in Vermont, in any case, participants are gradually moving toward a view that I have held all along which is that the contents of the portfolio, as Dr. French mentioned, have to be very carefully circumscribed. You have to say to people, "Here are the kinds of things that we can rate, given our criteria, our standards. Here are the kinds of things we either don't want to rate or can't rate." And the reason I want to stress that is because one of the kinds of products that raters in Vermont had difficulty with is non-verbal products. What do you do with a video tape? Well, depending on your criteria and your standards, your exemplars that you give children, a video tape may be perfectly appropriate, but for other sets of criteria, it's unusable, and if, for instances, what you are interested in is the ability to communicate mathematics, a non-verbal video often won't tell you much. Now, what this is leading up to, I think you all can see, is that whether or not a portfolio system is a better, truer gage of what LEP children can do, despite their limited proficiency in English, depends on what you say should go into a portfolio and what you say should not go into a portfolio depends entirely on what you want to say at the other end, what inferences you want to draw about it, about student performance. This is a big open question really right now. I don't think there is firm evidence that portfolio systems will, in general, be harder for LEP students, but that is my suspicion, and there is no evidence that, in general, they will prove better, in the sense of revealing more of what they can do despite their difficulties with English.

Now, some observations that are specifically about language and portfolios and LEP children in portfolios: It's very, very difficult to avoid the confounding of language and other skills when you do portfolios because of what goes in them. First of all, in the case of writing, that is obvious, but even in the case of mathematics, if you are going to do more than a traditional test, you want to find out what children can do, what they can explain, how they did things, and almost inevitably, you start drifting into a mix of whatever other things you want to measure and language.

In Vermont, math and writing were both assessed in grades 4 and 8 this past year, and the raters found that to be a very serious problem, even though in Vermont, as many of you probably know, there are virtually no children with limited English proficiency. It's a very homogeneous state, but even so, the raters often felt that children might be rated high in some cases because their math was good and, in other case, because their language was adept.

I want to make one more point about the confounding with language. That's not necessarily bad; it's a question of what it is that you are trying to measure and what it is you are trying to conclude. Many of you may be familiar with the new National Council of Teachers of Mathematics (NCTM) standards for curriculum and evaluation. Those standards are widely accepted, in name at least, in the education community, and they stress communication of language as one of the primary goals of instruction. An assessment that is designed to match the NCTM standards could not be designed to find out what they know but can't communicate. It is designed to find out what they know and can communicate, so there is no way of avoiding language.

I think that, in dealing with this, we are not getting into the controversial and stickier issues and, in dealing with them, I think it is necessary to separate technical from philosophical issues. Whether or not language is confounded with mathematics in a portfolio may be a technical issue. When you want language proficiency to be measured by a portfolio system, it is not a technical issue. It's a philosophical issue, and I'll give you some examples that are kind of silly because they are so extreme.

Many years ago I gave a Wexler Intelligence Test to an Israeli graduate student at Cornell whose English was fabulous. He had been studying in American schools since seventh grade. His father had been a diplomat. He was then a graduate student in sociology, but a few of the tests really threw him for a loop. One of them, called Digit Span, requires the reciting of even longer strings of numerical digits to the student, and the student is supposed to repeat them back. You see how long a span the student can remember, and then the student has to do it backwards, which is harder, and you are doing it against a stopwatch, which is very unpleasant. Well, this fellow really started to stumble. This is peculiar. So, just out of curiosity, since it was not for a formal evaluation but for practice, I switched to Hebrew, which I spoke, at that point, quite well because I used to live on a kibbutz, and immediately, his performance picked up. And, in that case, you would probably want to give that subtest in the person's native language, because you are trying to draw inferences about something that has nothing to do with language. It has to do with the ability to memorize spans and digits, but what if you are trying to measure, for example, a person's ability to write, or person's ability to explain solutions to mathematical problems, there's no choice. It has to be in some language, and what language should it be in? Well, that's a philosophical question.

The application of portfolios to LEP children underscores the difference between instructional and measurement goals. It's true, in general, for portfolio assessment, and again, this is a philosophical question, there might be a case where you prefer that a student's instruction be in English to give that student practice or might feel that student is not yet advanced enough in English that the assessment would be fair in English. There might be times where the reverse is true, depending on the situation, what you are trying to infer from children, what the particular children's abilities and goals are.

Finally, I will wrap this up, because we are running a little late, and I think there should be time for some discussion and argument. Using portfolios with LEP children, opposes some really substantial, practical constraints depending on the decisions you make about philosophical questions. Dr. Russell French raised the question of using raters who are fluent in the student's native language. Well, that might or might not be possible in a district such as Houston, where a large share of the population is LEP and almost every LEP kid in the district speaks Spanish as a native language. It is not possible, even remotely possible, in many other schools. In my neighborhood school, which is not even a high minority school, there are native speakers of Spanish, Portuguese, Japanese, Chinese, Swedish, Norwegian, and Hebrew, and there is not, to my knowledge, a single teacher in that school who speaks any of those languages fluently, let alone enough teachers who speak them fluently enough to test the reliability of scoring. The real world for most districts is that portfolios are going to be assessed by people who speak English and not anything else, and that raises very serious questions about how the portfolios ought to be run for LEP children and how they ought to be scored.

What potentials do portfolios have for LEP children? Personally, I think, in one sense, they could be a very big step forward, and here, I am speaking less about a technical view than a personal view. I think there has been a long and unpleasant history in the United States of giving children who have difficulty with school, for whatever reason, whether it be lack of facility in English or whatever, an even more boring diet of course work than regular kids get. This might be a big step away from that. Rather than taking children who have a little difficulty in math or difficulty in tracking the directions in English and saying, "you're going to do even more drill and more practice, until you are bored to tears" you say, "you're going to do some difficult work that actually makes you think and write, just like anybody else." Will these assessments tap abilities that some of these kids have that standardized tests might not? I think that it is very unlikely. I think that a portfolio assessment is inevitably going to, first of all, put more demands on children, especially LEP children, and second, bring to the fore some very difficult, not just technical but philosophical, issues about how LEP children are to be taught. I'll leave it at that.

A Political/Sociological Critique of Teacher Education Reforms: Evaluation of the Relation of Power and Knowledge[1]

Thomas S. Popkewitz
The University of Wisconsin-Madison

The last decade has seen a resurgence of interest in the problem of educational change. School reform is viewed as a mechanism to achieve economic revival, cultural transformation, national solidarity, and ethnic aspirations. An important part of the reforms concern the improvement of teaching and teacher education. The impetus for change has come from multiple sources: Federal and state governmental and philanthropic reports have focused on the quality of teaching, university curriculum, and student achievement. Legislation has increased the state's direct control over the policy and content of teacher education. A professional infrastructure has supported new programs and standards as ways to alter occupational practices, to increase teachers' remuneration, and to improve the quality of teaching. Central to the literature is a call for more educational research and professionalism among teachers.

Current reform practices should be viewed as an integral element of the events and structured arrangements of schooling. As a primary institution for establishing direction, purpose, and will in society, schooling ties polity, culture, economy, and the modern state to the cognitive and motivating patterns of the individual.[2] Educational reform does not merely transmit information on new practice. Defined as part of the social relations of schooling, reform can be considered a strategic site in which social regulation occurs and power relations are embodied.

It is within this context that I wish to explore the promise and limitations of evaluation in teacher education. The promise of evaluation is to understand the diverse issues and complexities that underlie the processes of reform; and to contribute to a more informed policy making. This focus is important to all who wish to promote intellectual integrity and social equality in schooling. The importance of evaluation is revealed through recent social theory and methodology which highlight the ways in which the categories, distinctions, and differences produced in social research establish social interests and power relations (see, e.g., Bourdieu, 1984; Cherryholmes, 1988; Clifford, 1988). Since evaluations are typically commissioned by those with power but in the name of a common good, the social values and relations that underlie research are important to consider.

Three themes in teacher education evaluation are considered. The first two themes are a cautionary tale about evaluation -- evaluation is produced in social fields in which people vie for authority. These themes focus on the power relations "carried" as the problems and strategies of evaluation are constructed.

1. Evaluation needs to consider issues of social production and the social realization of policy. This entails two dimensions. Evaluation is a state strategy to produce social amelioration. State is used as a theoretical category to explore how the larger concerns of social regulation and steering of institutions are carried into the daily life and practices of schooling and teacher education. The strategies applied, the categories and distinctions that construct the reforms, and the social contexts of teacher education and schooling interact to produce social values and power relations.

2. The distinctions, categories, and differences embodied in evaluation are not neutral terms to describe events; they are modes of presentation and styles of reasoning that construct the subject; tying discourse to issues of power. Words which have currency in evaluation (e.g., measurement, assessment, professionalization, empowerment, and site-based management) have no fixed and unyielding meaning but are constructed in historical contexts and institutional settings. We must take into account the social contexts in which the words are used; entertaining a skepticism about practices that offer to make the world better.

3. The third theme pursues a central issue about the purpose of evaluation. It argues that evaluation has a policy clarification purpose. It can help to illuminate the tensions, contradictions, and ambiguities that underlie the realization of educational reform; it does not tell us what policy is most efficient or useful. This may seem obvious. Reforms respond to perceived issues and problems that, at face value, are not clearly defined and do not have linear outcomes. Of the deepest value to the public debates around which schooling in a democracy (and of importance not only to policy makers) is an understanding of the strains and tensions found in the relations in school arenas. Evaluation, at its most productive sense, considers the tensions, struggles, and ambiguities as social practices relate to social goals. Further, the reform priorities of teacher education are indelibly tied to social, cultural, and economic conditions; these cannot be lost in the methodologies of evaluation.

Recent studies of teacher education and teaching will provide illustrations of the relation of reform, knowledge, and power.

I. Social Production and Social Reception of Policy

At least two related issues are central in evaluation. One is the relation of evaluation to state planning. Second, is the realization of reform in social fields that "carry" values and interests that are not necessarily those of the program planners. As a result, the strategies and procedures of reforms maintain social values that should be scrutinized.

State Policy, Policing, and Evaluation

Evaluation is a part of state regulation, monitoring and steering. In this sense, policy and policing are epistemologically related; policing, in its French and German origin, refers to the specific techniques by which government, in the framework of the state, enabled individuals to be useful to society (Foucault, 1988, p. 154). Older forms of evaluation involved political arithmetic or statistics in which the state collected demographic and other data to steer reform policies during the formation of the modern state (see, e.g., Haskell, 1984). More recently, evaluation is intended to provide public accountability for different and sometimes contradictory reform strategies (such as to introduce standards that make a citizenry that is more productive in an arena of increased international competitiveness while, at the same time, to provide a humanism that allows for cultural and social diversity).

On the surface, the current situation of evaluation has a particular historical character. Evaluation is considered necessary for decision making and accountability. But to understand this situation, we need to think relationally about the state, local community, and schooling. In part, evaluation emerges as a professional field to respond to increased governmental involvement in the educational sector following World War II.[3] Further, the particular form that evaluation took in the United States involved particular social constellations. There is a long standing commitment to local governance, individual school improvement, and university autonomy in professional education. This commitment to the local and the "individual" occurs, as Weiler (1990) argues, as part of state formations in which accountability and steering are of great importance. It is related to the need for competentory **legitimation**. United States evaluation strategies should also be seen as maintaining historically derived commitments to define change through individual practices (Meyer, 1987; Popkewitz, 1991).

While we often value the individual and local governance over state rule (e.g., school site management), we cannot disregard the societal purposes that are part of the normative construction of state

299

agendas. We need to understand how social goals, articulated through state policies, become reconstituted as they are realized in the institutional practices (Lundgren, 1990). Thus, evaluation has a policing quality in the modern state, whether we see it as part of the noble intent and desire of those who seek to improve school or as part of the darker side of social regulation.

My reason for starting with this assertion is neither to demean the effort of state actions nor to pose an anarchist view of the social processes of schooling. Rather, my intention is to remind the reader that evaluation is not merely a strategy that "objectively" describes outcomes of educational practices. This becomes more crucial in the United States where there is historical anesthesia toward school as a state institution; social/political values are hidden in research paradigms of teacher education that emphasize change as individual, teaching as a problem of psychological motivation, and as sociology that is centered on organizational efficiency rather than social relations of schooling (see Popkewitz, 1984, 1991). The problem of evaluation must be positioned within educational fields that include studies of the power relations which, at root, contribute to the processes of social production, regulation, and the creation of human capabilities. (See Bourdieu, 1989 for a discussion of the problems in the social field of intellectuals.)

Reform in Social Fields

With the state as a central actor, the problem of evaluation is constructed within particular social fields and power relations. The questions, conceptual schemes and "tools" of teacher education contain assumptions, debates, and implications to how questions are framed and solutions legitimated.

The public discussions in the United States give attention to the changing international character of economic relations, the redesigning of national priorities in schooling and the need to maintain greater cultural strength through the socialization processes of schooling. This new nationalism stresses the country's international competitiveness while focusing on local flexibility and semi-autonomy -- with national standards by which to judge local attainment. In contrast to the 1960s school reform efforts in the United States, which made curriculum issues focal, the current reforms give attention to teacher quality, standards of work and professional education. Market metaphors (i.e., choice) are combined with those of outputs (accountability) in the current reforms, looking at output measures such as SAT scores to determine progress.

Teacher education has been a centerpiece of these reforms. Research has focused on the qualities of a good teacher and sought to emphasize those qualities in programs. Greater emphasis is sought on relating pedagogical questions of teaching to the cognitive disciplines in which teachers work. Research on teaching of mathematics, technology, and science is sponsored by the U.S. government and foundations. Priority is given to relating psychological paradigms for translating disciplinary knowledge into school subjects; often with professionalism in other disciplines fighting for the legitimacy of their subjects through obtaining research and development funds. Model programs are tried; professional schools established for training teachers, and greater attention is given to whom comes into teacher education programs (see, e.g., Holmes Group, 1986, 1990). Criteria for certification and credentialing have become revised by state governments to reflect economic goals of teaching mathematics and technology and social goals related to making schools responsive to the diverse populations in the United States. There is continual reference to the amount of time teachers spend doing administrative work and routinized activities, such as collecting and grading papers or responding to central office requests. The language of reform seeks to produce a more professional teacher corps that has increased status, responsibility, and financial reward.

We can evaluate the reforms at different layers of understanding and interpretation. One is a tendency to consider the behavioral and organizational conditions of reform: Are teacher education programs revising course syllabi? Are programs giving students adequate time in practical experiences? Are students having opportunities to work with diverse populations? The notion of a National Report Card issued by the Department of Education follows this line of thinking. At a different level of evaluation, there is a focus on the conditions in which student teachers work: Do student teachers, for example, have time to share ideas and to attend professional conferences? Notions of standards, collegiality, professionalization, and community appear to frame these questions and the successes of reform. There is greater reference given teachers, as they are to be more autonomous and responsible in their work place.

The approaches follow what is often done in school evaluation and is a standard of the larger paradigm of educational research from which it is drawn.[4] The properties of evaluation are viewed as having conceptually distinct and ordered qualities that could be controlled and manipulated toward some desired end, with utility of practice given greatest value. Evaluations assume that organizational activities can be modified by more efficient management, and the results of a planned change defined and measured against cost (monetary and social). A result tends to be random collection of data about surface (observable and measurable) qualities of teaching and teacher education.

In this paper, I want to argue for an approach to evaluation that focuses on the social patterns and conceptions of knowledge that order these patterns and conceptions. Our concern is how reform practices organize and give value to certain types of social relations and, at the same time, structure out of consideration other possibilities for education. This layer of analysis enables us to consider the organizational and perceptual characteristics related to a reform but provide conceptual ways of understanding the assumptions, implications, and consequences of social practices.[5] Examples from an elementary school and teacher education evaluations provide an illustration of the social complexities with which evaluators must grapple. Then I will proceed to reconsider the theoretical issues that underlie these evaluations.

The Social Complexities of An Alternative Certification Program: An Example

One of the major changes in teacher education has been the introduction of alterative certification programs. These programs provide ways in which college graduates in non-education majors can teach in critical areas without going through a regular teacher education program. The particular one that we examined was national in scope and sought to bring into rural and urban schools students who graduated from liberal arts colleges but who wished to make a commitment to teaching for at least two years.[6] About 500 recent graduates volunteered and attended an eight-week summer session to prepare them for teaching that fall. Once in their teaching sites, they would work as regular teachers while obtaining certification. The alternative certification program fills an important niche in teacher education: directly recruiting and training teachers in areas where there is a severe shortage. In addition, the corps of teachers are people who are well-educated in their particular field.

As part of the evaluation of this program, we sought to understand what was being taught to the first-year teachers as they entered schools and how these efforts relate to existing social relations that define urban and rural schools. The evaluation, then, was to consider how the "components" of the reform program was realized in its social contexts: exploring the ways in which discursive patterns and the institutional practices that structured the linguistic and classroom practices of the new teacher changes. We did a survey of mentor teachers and administrators (and found that they liked the performance of the first-year teachers); but also sought to measure the ways in which teachers, administrators, and first year teachers structured the problems and tasks of teaching and the conceptions that they held of knowledge, teaching, children, and community. With these statistical data were systematic observations and interviews in each of the five regions in which the new teachers were placed. The field data were collected throughout the summer institute and school year. Some tentative summaries of these relations can be identified here:

- There was a shift in the perceptions of the first-year teachers from an idealism that saw a teacher as a missionary to one who had to learn classroom management and control for success. Goals were revised. Teaching was viewed as part teaching textbook content and part motivator of children who the teachers saw as having little self-esteem.[7]

- The textbook and testing became the center of curriculum. This responded to a variety of "control" factors. It was to control content when absenteeism and movement among families provided very little physical continuity in classrooms. It was a control mechanism in school districts where there was little money for supplies and books. Social control was also a characteristic. The strong regimentation associated with textbook teaching and continual testing was to instill discipline. This was through the information conveyed and through the rituals of social interactions applied. While clearly it did not work, the ritual of practice created a sense of order to the curriculum and social patterns in classrooms.[8]

- The first year teachers developed an "educational" language that shifted attention away from the social conditions that impacted on their teaching and legitimated ongoing practices of the school. This language was not technical but based on rules and standards of reasoning that is associated with schooling. Problems were made into those of the psychology of the individual or the pathology of the community. The novice teachers talked about setting educational objectives as central to their professional role or about the need to develop self-esteem among the students before they could learn properly.

- The language of management and psychology were used to evaluate the competence of teachers themselves. A rural principal talked about the first year teacher having a difficult time because of the poor motivation of the children and the lack of management skills. There was a tie of competence to control which is then related to whom the children were -- from poor families and of color.

- These languages about the problems and solution of educational problems recast practices so as to make the problem of reform as one of learning some "proper" content, individual initiative and psychological characteristics rather than of structural concerns. The language coded the racial and economic distinctions within the district. In rural areas, schools were internally segregated by race, with people of color tracked in the lower class. In urban settings, segregation was the order of the day.

- There was an emphasis on practice as learning about teaching. There was talk about "the real difference between "theory" and "practice." This set up a dualism in which what teachers did in schools was given value. This, in turn, gave value to existing social patterns of school practices; thus legitimated, as an unintended consequence, social inequities carried in the day-to-day instruction.

- The alternative certification program intersected with other school reforms in a manner that decontextualizes and reformulates social issues into administrative ones. Mentoring systems, an important reform to help first year teachers, tended to focus on advice about how to work in a bureaucracy, such as getting the duplication needed or what networks exist within the school to get supplies, or how to plan with objectives. The discourse structured out consideration of what was selected to teach, and the social/cultural contexts in which teaching was realized.

- State definitions for social amelioration were reformulated through the social processes in schooling. State definitions of children in need of special help define those who come to school as special "populations" in need of remediation. While not policymakers' intent, labels, such as a "Chapter One School" were used to consider the pathological character of the school's students.

- The official categories created a history for the schools; the official categories defined the enterprises from distinctions drawn from policies about social amelioration and regulation. The official language about reform co-existed with teacher discourses about the management of classrooms and psychologies of children's competence that gave focus to individuality as a "core" assumption to define teachers, school, and student competence or failure. The reform discourses and the teacher "knowledge" were mutually sustaining in these assumptions about the world of schooling. The professional classification filtered out consideration of the social complexities of the situations of schooling, in some cases redefining the rich histories of schools as a community institution. In their place were references to the particularities of individuals as part of statistic/aggregates that defined them as economically, culturally, and socially poor. The family/community were symbolically represented as populations which have no history or sociality except as part of the aggregate used to group them as in "need." The use of phrases such as "schools in transition" also provided a language that reconceptualizes and reformulates who the children are and the tasks of the school.

In each state, the national teacher recruitment program had to respond to state governmental requirements for certification. Alternative certification programs administered by states reinforced this valuing of the immediate and useful in teaching. Certification courses focused on methods of mainstreaming, for example, without dealing with the social and political debates that underlie the pedagogical approaches. In some states, first-year teachers had to spend 150 hours writing their lesson objectives and reports of meetings held with their mentors as part of the certification requirement. Reflection, another slogan of teacher education reforms, was sociologically restricted to writing objectives about upcoming lessons.

The study of the alternative program provides a way to consider how an innovative program is located into institutional patterns and discourses about education. An alchemy occurred as the public rhetoric about reform passed into the social space of schooling. Institutionalized practices and professional discourses shaped and framed boundaries by which reform give reference to economic issues and cultural debates. The social density and the mobilization of discourse constitute and express an ordering of the world that teachers inhabited. We can think of the categorical and syntactical procedures in schools as establishing hierarchies, relations, and values. I will return to this issue in the following sections, after I provide a second example.

School Reform, Classroom Cultures, and Social Differentiation: An Example

In this section, I want to explore how pedagogical practices themselves carry power relations that have cultural, social, and political implications. In a study of an elementary school reform (Popkewitz, et al, 1982), for example, we explored how six schools in different parts of the country used a particular program called Individually Guided Instruction. As before, the focus was on the classroom dynamics and social relations in which the program was realized.

The initial expectation was to consider variations of implementation of the reform as all schools incorporated the same organization patterns, used the same curriculum, and had the same technical languages and numonics for talking about expectations and experiences -- IGE, ICC, multi-unit schools, differentiated staffing, and so on.[9] But rather than a common school, the schools we visited were different in their cultural organization of teaching and learning. These differences related to a number of different social phenomena -- education and occupation of parents, income, gender, race, and religion were intertwined in the productions of accomplishment and competence defined in schools. Thus, while students in different schools might use similar textbooks or participate in similar lessons about map reading, the cultural messages were different in the schools that we studied. How the maps or books were treated as a "learning" experience, the role of the student and teacher in deciphering the maps and books, and the authority to base the knowledge about people and places were very different. These social distinctions and differentiations in schooling need to be considered as part of the accounting of schooling itself.

We called these different social conditions of schools Constructive, Technical, and Illusory schooling. What is important here is that the conversations in the daily instruction were differences in the expectations and demands of teachers and students. Some stressed gaining information, others placed value on autonomy, responsibility, and the tentativeness of knowledge itself. A third went through all the motions of teaching and learning, but there was no carry through. The teaching had to do with trying to morally uplift the students through references to the students' background and the importing of other values. At a different level, the classroom discussions "told" students about how they should act, talk, speak, and think about themselves and the world in which they live. Implicit in the instruction was political theories about citizen involvement (passive or active) and the representation of what is good/bad or possible and not possible.

The different cultural messages carried in the schools had little to do with traditional notions of teacher competency; such as years teaching, classroom climate, levels of education, or classroom management techniques. The differences involved a complex relation between social/economic and cultural conditions in the community, that included, in some instances, religious beliefs and institutions, and professional discursive practices within the school. The constructive school was located in a professional community in which students brought to school certain expectations and teachers maintained perceptions about what is legitimate for students to do. In the Illusory schools, expectations and demands were built on the welfare situation, family life, and work horizons posited in the community.

Rather than a common school, there were different types of schooling for different children that had little to do with the formal criteria of achievement or competence in the relation of teaching and learning. These expectations and demands were established through the interactional patterns and cognitive structures that organized everyday life of teachers and students. In turn, these relations were reformulated into a school language of efficiency and psychology of the individual.

In both evaluations, the inquiry about performance, competence and "outcomes" went beyond the formal goals and policies of the projects to consider the power relations that were embodied in its social patterns. In both instances, the evaluations were conceptually driven to relate knowledge and institutional patterns to power. I would like to explore further this notion of power in the next section by considering discourses of research and evaluation about teaching and teacher education. Here, I will argue that evaluation needs to consider the categories and rules of speech about schooling and teacher education as forms of social regulation.

II. Styles of Reasoning and Constructing the Subject: Discourse and Power

In this section, I focus on a particular concept of power that was implicit in the previous discussion; one that explores the standards of reasoning, ways of thinking and rules of truth that underlie teacher education and school reform. The current reform efforts have a mode of presentation and styles of reasoning that are not only "telling" stories about schooling, teachers, and teacher education but also constructing the subject itself; establishing value and authority about the ways in which we define what is good, legitimate, and plausible about schooling and teaching. We can think of a constant litany of words in current educational reform; among them, professionalism, teachers' thought, content knowledge, codifiable knowledge and knowledge base, empowerment, reflection, teacher efficiency and practice. The words are not free floating words that have unfixed and unyielding meanings over time, but assume a particular nexus of relations, hierarchy, and value as teacher education relates with schooling. The distinctions, categories, and difference establish a cognitive structure about "Others" (students, parents, and community) and of "self"; the teacher and teaching.

My use of cognitive structures, however, has little to do with the current interest in cognitive psychology which defines the mind as independent of social and historical circumstances. My discussion of "structures" is drawn from the sociology of knowledge. In particular, I am concerned with how reforms embody a particular form of consciousness about schooling, teaching, and teacher education. As do Berger, Berger and Kellner (1973), I consider the resulting styles of thought and perception associated with teacher education as a particular type of consciousness about the world that is tied to particular institutions and institutional processes.

Evaluation methodologies need to consider language as an instrument of action and power. Under the guise of methodological distinctions, evaluation strategies establish particular sets of linguistic practices as dominant and legitimate. This language formation, however, occurs in specific social and political conditions. Presupposed are certain forms of cognition and belief about competence and performance. Agreeing with Bourdieu (1991), we can think of language as a form of symbolic power or violence; in which interests of particular groups are established through symbolic exchanges rather than imposing dominance through brute force. The symbolic violence occurs as the communicative exchanges are made to seem to rest on a foundation of shared belief, but the hierarchies of distinctions and differentiations fix values in a manner "that even those who benefit least from the exercise of power participate, to some extent, in their own subjections" (p. 23). The nature of symbolic power, Bourdieu continues, is that it presupposes an active complicity as the distinctions that legitimate, in our case, the calls for reform become the belief about salvation for those who seek to redress their oppressions.

Concepts and categories have two sides (Smith, 1990). There is the surface in which the concept and category abstract form, and express social relations. There is also an underside of social relations in which the concept or category arises. Reform concepts such as "individualization," "empowerment," "community," and "participation," -- part of the common sense in teacher educational reforms -- have meaning to the extent that the distinctions that they make are already apparent in the structure of their actual social relations. People grasp them as particular forms of ordering their practical activities. It is this underside of language in history that I believe is an important element to the ways evaluation is conceptualized.

The Language of Reform as a Structuring of Social Relations

Examining the structuring principles of a language about reform can be pursued through the "use" of the words, professionalization and professionalism. The current teacher education reform movement, for example, makes professionalization and school improvement as goals of policy in the current educational reforms. These words appear in various ways in the reforms of the United States, Iceland, Australia, Sweden, Spain, among other nations. The words, however, do not have an absolute character that refers to basic ideas or conditions of schooling. Words do not represent reality, they are part of its creation, sustenance, and renewal. Words refer to concepts that change in relation to their position with other words and in relation to the social conditions in which they are used. To examine its use is to understand how power can operate through the effects of discourse.

It is apparent as we examine the historical literature that there is no understanding of profession in any universal manner. There are important differences between the Anglo-American and European continental traditions of professions. In part, these differences have to do with the different forms of state developments in relation to certain middle class and elite occupations. In France, for example, professions were sponsored by state agendas and tied to the development of state agendas through Napoleonic reforms. British and American professional development occurred through changes in civil society but quickly established relations with the state, including sponsorship of the nascent social sciences. In many European countries, such as Germany and Spain, there was no word similar to that of the Anglo-Saxon word "profession." In Germany, the "bildungburger" refers to an educated class and the word "professional" has only been incorporated in relation of academic discourses (Kocka, 1990). In recent years, however, the Anglo-Saxon word "profession" has been brought into the language of many continental countries to describe the social formations of work within the middle class and increase importance of expertise in the process of production/reproduction.

To incorporate the Anglo-American conception of profession imposes an implicit interpretive "lens" about knowledge, occupations, and state. At a social level, there is an assumption of an occupation controlling a market of liberal theory about society as based on individual social contracts, and entrepreneurial relations are given privilege (Collins, 1990). At an epistemological level, the individualism is based on a particular managerial conception of knowledge; social phenomena and individual "development" can be rationally and hierarchical order to provide for social betterment. These assumptions (social and epistemological) are found in much of the discussion about American professions in which ideal types of disinterested occupations are offered that are separate from the state. Autonomy, technical knowledge, occupational control of entry, remuneration, and high ethics dominate this recounting.

The ideal types, however, have little basis in fact; they ignore the political struggles, debates and compromises involved in the formation of the professions. Nor do these types account for the ways in which modern professions become a part of the social regulation and governance structures of the modern state (see Burrage & Torstendahl, 1990). The major purpose of these ideal types is as legitimating strategies for maintaining cultural and social authority, not for analytical purposes.

The "ideal type" of profession has assumed importance in the United States as an issue of school improvement. As an ideological stance, it seems difficult to argue against; teachers should participate in their work with autonomy, integrity, and responsibility. At this level, the slogan is important to a reconstruction of schooling. But to talk about professionalism, integrity, and responsibility without focusing on the content of the participation and the structural relations that shape it, is to lose site of the historicity of our practices.

309

"Professionalism" has been an aspect of reform in U.S. education since the early 19th century; it referred to two different layers in the formation of the occupation. Professionalism was a slogan for those "at the top," including administrators and university professors. This occurred with the specialization in university training and the development of new disciplines of educational sciences (as part of the planning and evaluation of schooling) (Powell, 1980). This strata of the occupation tended to be male-dominated and better paid than teachers (typically women); administrators and professors enjoyed degrees of responsibility in the conceptualization and organization of their work conditions. The organization of higher education and research in education provided routes for the occupational mobility of men. Further, much of the research and evaluation schemes for schooling relegated teachers to an ancillary status and focused on assessment of large groups of students and indicators of qualities of the entire school system. The administration also contained elements of social regulation for those at the bottom of the occupational ladder -- usually in the name of school improvement or effectiveness.

At the bottom were teachers. Many of the reforms of the late nineteenth century made teaching more bureaucratic in the name of professionalization (see, e.g., Mattingly, 1987). Standardized hiring practices, uniform curriculum policies, and teacher evaluation practices eroded spheres of teacher autonomy and responsibility through an increased rationalization of school organization and didactics. At no time in the history of modern mass schooling have the working conditions of teachers in the United States provided the opportunity to reflect about their situation in a sustained manner. Teacher education has been pragmatic and fragmented; it devalues an intellectual focus.

In fact, the words of profession have taken on meanings which tie the regulatory values found in U.S. schools to the concept of bureaucracy that the new reforms were instituted to change. Their assumptions about and categorizations of social phenomena contain a classical Weberian formulation of bureaucracy. Social reality becomes one-dimensional to include particular categories of people without considering the substantive quality of the resulting interactions. Its products are seen in relation to utilitarian values. Community consensus and participation are based on administrative criteria that define people as interest groups with homogeneous values; the conflict and debate about purposes that cross the lines of the designated actors are structured out of consideration. Internally, school processes are seen as orderly; elements can be placed into proper and, thereby, manageable places. Each component has a self-contained unit with a specific relationship to other elements. Action is seen in relation to an abstract frame of reference that is divorced from the specified, complex tasks of teaching. Denied is a sense of the history and the power relations involved in the formation of schooling (for a different conception in teacher education, see Tom 1990).

These trends continue in current reform efforts, linking professionalism to school improvement. A study of three school districts (Popkewitz & Lind, 1989) involved in an effort to increase teachers' professionalism indicates this very clearly. The reform strategies increased the teachers' work load and the level of monitoring of teacher practices. Evaluation was to provide "evidence" of teacher accountability and more rational approaches toward school improvement. It was assumed that there is a direct relation between the knowledge of evaluation and specific practices and actions. Teacher evaluations valued instrumental and procedural concerns and devalued the craft and expressive elements of teaching -- those elements that have gender implications.

In one of the sessions in which teachers were considering evaluation approaches, they brought in an outside expert who owned a commercial company selling evaluations. He argued that the evaluations were created because teachers did not want to do them, "they wear their heart on their sleeves," he said. The proposed evaluations would make the tasks of improving teaching scientific and objective. Teachers would go into another's classroom to observe and record by checking off words that described the classroom: well-managed, enthusiastic, directive, scholarly. On examining the words, it was clear that the categories described tightly controlled classrooms and said nothing about what was taught. Further, behaviors such as "expressive" were devalued through the rating system applied. This subtle emphasis and deemphasis can be related to issues of gender itself.

This brings us to two issues which can be summarized briefly. One is the complexity of reform. To solely consider the organizational and behavioral elements of reform is to obscure the manner in which knowledge about reform, teacher thinking, and school practices interrelate as an effect of power. Further, the discussion of teacher improvement in the United States carries certain historical assumptions about social relations and progress. Two, and which I will explore further in the next section, the categories of research and evaluation do interrelate to provide boundaries that structure out certain possibilities and legitimate others through the discourse that is constructed about school change.

The Ordering of Populations and Issues of Regulation

Part of the cognitive structures that represent **and** produce a "common sense" about teacher education is the defining of people as "populations." The concepts and procedures of social science and bureaucracy inscribe people as having discrete attributes by eliciting information and data or by establishing categories and codes for observation and recoding (Smith, 1990). The concepts make it seem that what is out there is what actually happens. For example, the demographers' use of sex as a category counts sexiness not as a matter of speculation but involves gender presuppositions about the actual everyday practices that are generated as a feature of the social organization in which the demographer works.

Let me take as an example from transcripts from recent interviews of first-year teachers' working in the Los Angeles area. Many of the teachers came from private, elite universities and saw the teaching experience as "giving back" to society for the privilege that they has had. Much of the curriculum, however, was designed around textbooks and testing.

One new teacher discussed the pressure for tests and school standards as related to the social background of the children who came to school. The teacher of Spanish thought that the school places requirements on children that they do not need, such as learning Spanish.

> Students need English. To students that need to be able to write simple sentences in English. To students that need to be able to carry on a conversation without saying "ain't." Or "got none" or any of that, that there are a lot of requirements being put on the students by the school that the teachers don't even have anything to do with, that we have to teach that these students really, they need to know it, but not as much as they need to know other stuff. Not every one of my students needs to know Spanish, but every one of my students needs better English skills. Desperately.

The teacher places this pressure on certain requirements and the students "needs" in relation to the African-American homes and community. She said:

> I didn't realize the background my kids came from. I didn't realize, I didn't realize that my background was where I had a safe place where I could go home and study as much as I wanted whereas these kids, they're lucky if they can sleep at home, let alone do anything else. Even watch TV. All they do when they're home is have their parents yell at them and have their parents blow smoke in their face from their cigarettes and things like that. They can't study. And the school is just so disruptive, like is going on now, there's so much pressure not to learn, there's so much pressure not to do what's expected of you that the best that most of these kids can hope for is to get through here without being permanently scarred.

Populational characteristics inscribed on these students are: crime figures, teenage pregnancy, single and extended family homes, and so on are made into representations of the "Other" who are in need remediation; the "At-Risk" children of the school. The language of populations objectifies the immediate and social relations of these students and constitutes and expresses them as separate and distinct from the subjectivities and "real world" of social relations in which they live. The social organization of reading the factual accounts "inserts" categorical and syntactic procedures into the actuality of education; thus establishing a normalcy to schooling based on pathological distinctions.

In teacher education, processes of learning how to teach inscribe the "population" distinctions into the attributes of students and teaching. The ongoing practices of different professionals in particular sites and across sites are coordinated and standardized through languages of "at risk," "learning disabilities," and the label of "Chapter One" schools. In these objectivifications of schooling, teachers learn how to relate to members of their own professions and to those of others, and they learn how to talk **to** students and how to talk to students so as to be able to talk about students that responds to structural and power relations. The psychological categories of affective and cognitive attributes and definitions of achievement also posit not only ways to talk about schooling, but linguistically conelates to strategies of lesson planning and teacher reflectiveness.

The constructing teacher education as an object of scrutiny also entails power relations. My concern with power is with its *effects* as it circulates through institutional practices and the discourses of daily life. Here, the work of Michel Foucault is most helpful in understanding how structures of thought are practices that construct the objects of the world rather than represent those objects. This concept of power, Foucault argues, is embedded in the governing systems of order, appropriation, and exclusion by which subjectivities are constructed and social life is formed. This occurs at multiple layers of daily life, from the organization of institutions to the self-discipline and regularization of the perceptions and experiences according to which individuals act. Power is embodied in the ways that individuals construct boundaries for themselves, define categories of good/bad, and envision possibilities. The effects of power are in the production of desire, dispositions, and sensitivities. Power, in this latter sense, is intricately bound to the rules, standards, and styles of reasoning by which individuals speak, think, and act in producing their everyday world (see Foucault, 1988; also Dreyfus & Rabinow, 1983, Noujain, 1987; Rajchman, 1985).[10]

Here I am countering the folk-wisdom of research on teaching which says that teachers do not have a technical language and therefore are not professional. When discourse is examined rather than the use of words, teachers express a language in which certain categories, distinctions, and differentiations are made that relate to institutional practices. Teachers seem to have certain standards and rules of speech that establish an occupational "identity"; and that language distinctions between what is to be spoken about as schooling and what is not considered as legitimate speech.

III. Tensions in the Relation of State and Teacher Education: A Problematic for Evaluation[11]

The two previous themes focused on the historical limitations and contradictions of evaluation; I would now like to focus on a different element in the problematic of evaluation: the promise of evaluation is not in foretelling what is to be done, but in understanding the tensions, dilemmas, and ambiguities that underlie the current transformations. Reform is not an object that can be installed or that has essential properties to be discovered. The processes of reform are dynamic not static or linear; therefore, evaluation requires a scrutiny of the relations that occur in its setting. Further, it will be argued that evaluation is inevitably about the past rather than the present or future. Its contribution to policy making is through an illumination of the dilemmas, tensions, and contradictions of how things have worked.

Evaluation is not a problem of school improvement as it is traditionally assumed. Nor is it a problem of "testing" or "verifying" state practices as though there is a consensus of values or a standard notion of progress. The problems of schooling are of a different order, one which evaluations can help to illuminate, but will not solve. At best, evaluations can provide a greater understanding of the tensions, struggles, and dilemmas that underlie efforts toward social improvement; the very moral, political, and cultural complexity of schooling as a social endeavor makes the search for precision and certainty as a chimera.

Let me conclude this section by saying that I do not believe evaluation can provide a direct plan of the present or of the future. Evaluation is inevitably about the past. Epistemologically, evaluation and social science are dialects of language and involve interpretations of what has happened. While we like to think of our generalizations as present-centered, we are constrained by the constructions of narratives that occur after the events. The language that is used in schooling that built on the outcomes of past struggles. There is also a political question here. When we adopt a belief that knowledge is about prediction and administration, we allow science and its relation to the empirical world to move into the realm of ideology and social control. The rituals of science and evaluation become a rhetorical form whose purpose is to convince others that what is being done to them is in their best interest.

I say this because I can find no evidence that social science and school evaluation have anything to say, **qua** science, about the future. They do offer methods for understanding the boundaries that exist in the past and the dilemmas that are embedded in those arrangements. This is not to say that science cannot help us in the choices we make, but it is often in a negative voice. To borrow partially from Karl Popper, science (and evaluation) do not verify but refute. They can help us understand what choices to make, such as in eliminating fluorins, controlling the deforestation in the Amazon, or limiting the use of intelligence testing. But -- in the policy arena -- the findings of science are part of a public debate that rarely concerns evidence alone. The determination of futures is no longer reserved for particular elites and experts who claim a sacred knowledge.

Before ending this discussion about past, futures, and evaluation, there is an important caveat. Evaluation is about the future in an indirect way. The categories of evaluation organize phenomena in a manner that sensitizes us toward certain possibilities and, at the same time, filters out others. Implicit in practices, then, are ways in which people are to challenge the world and locate themselves in its ongoing relations. This is what Steiner Kvale (1990) has made clear in his discussion of evaluation as a knowledge and constituting practice which produces a censorship of meaning. To focus on the problem of tension and ambiguities is not to remove the necessity of collecting demographic or achievement data, but to make the collection of data responsive to the central issues posed in the evaluation. The role of evaluation and evaluators in the ongoing construction of the world is one of continual scrutiny.

IV. The Enlightenment Project as a Problematic for Evaluation

What I would like to propose here is a reconsideration of an old Western European commitment to the Enlightenment, although recasting it, as a major purpose of evaluation. By the word "Enlightenment" I mean a view in which people are assumed to have competence and responsibility for the governance of their own lives. Schools, in this context, are an important educative institution in the realization of that goal. The major strands of twentieth century European and U.S. philosophy, sociology of knowledge, cultural studies, and literary analysis refocus the problem of the Enlightenment on the boundaries to our existence, the ambiguity of knowledge, and the fundamental relationship of social practices, power, and knowledge. The redefining of Enlightenment project becomes bound with a recognition that there are multiple truth claims, that these truth claims are historically bound and emerge from the social struggles and tensions of a world in which we live, and that the production of human possibilities always contains contradiction. The question of evaluation is how schools of education and professional programs reflect and sustain these commitments.

Posing of a curriculum entails certain general and seemingly transcendent values that we wish to maintain in schooling. In European traditions, these values relate notions of democracy to reason. Curriculum supposes philosophical assumptions that reason and rationality can help improve social conditions; political assumptions about the relation and responsibilities of people and institutions; and cultural assumptions about the central values and patterns that should give direction to social affairs. Yet contemporary scholarship makes us aware that however noble our hopes, a curriculum is a socially constructed and politically bound practice. At all times, our language and social practices in schools are precarious and limited, containing contradictions. As we engage in the tasks of constructing and realizing a curriculum, what are defined as possibilities are also prisons.

The problem of evaluation, then, is not merely that of school improvement or decision making, but of the conditions under which and the manner in which the knowledge of schooling is produced in teacher education. Evaluation needs to focus not only on fundamental assumptions about the purposes of schooling that underlie practice, but, as significantly, on issues about the relation of individuals to society which exist in the constructions of pedagogy. Evaluation should promote a discourse about education which examines the ways schools illuminate or obscure the social conditions in which people live. While I recognize the difficulty of conceptualizing such assessments, I believe that attention should be continually directed to what is most important through carefully considering the conceptualization by which data are defined and collected.

The complex and profound problem of curriculum can be expressed as a conflict between the hope we place in schooling and what happens as people seek to create, sustain, and renew the conditions of their world (see Lundgren, 1983). The history of curriculum is one in which theories are never realized in the manner they are intended. As theories are put into social practice, there are always unintended, unanticipated, and unwilled consequences as theories are put into social practice.

Here, I am taking up the theme of the social construction of knowledge that is so prominent in social theory and philosophy (e.g., Rorty, 1989; Giddens, 1987). My interest is to consider a socially constructed knowledge as a strategy for the construction and evaluation of school curriculum. This "turn" to constructivism, however, is not a psychological one that focuses on how students mediate a given knowledge. It is sociological, historical, and linguistically based. Science, mathematics, literature, and art are to be considered as social fields which they are -- multiple and competing ways of thinking and acting toward the world. These paradigmatic endeavors are struggles for authority about what is legitimate truth.

Evaluation needs to consider whether the selection of school knowledge pays attention to the variety, debates, and tensions that exist in how people come to know and interpret their world. It is also to consider the relations of power/knowledge in those formulations of teaching and learning.

The problem of school learning and evaluation is to consider how students come to grips with the human constructions of knowledge--its fragility, tentativeness, skepticism and change. It is not to correct misrepresentations as the psychologists of education would have us believe; but to consider the variety of representations that exist and how systems of thought are practices that shape and fashion social, cultural, and political worlds. It is to recognize various dialects in schooling as tribal and partial. Whether reforms focus on teaching science or on the heritages of various peoples who live within the United States, the evaluation of practice should direct attention to the types of reasoning developed, and the means by which both a trust and a healthly skepticism toward the world can be accomplished.

We can also think of the everyday world of schooling as having differentiations that are produced through the social patterns of school life. The patterns of conversation and the practices of teaching are not common to all; they contain multiple layers of meaning and interpretation. In back of the rituals of a common institution are social differentiations: not all children are taught the same things nor are the dispositions, sensitivities, and awareness common across social groupings. In light of this, we need to ask: What knowledge is to be transmitted to whom? What are the different cultural and social messages transmitted in classroom interactions?

How these problems are represented within our teacher education programs needs to be the problem of evaluation. Methodologies need to be constructed that give attention to how teacher education interrelates with schooling. This task cannot be defined technically, such as whether one uses qualitative or quantitative (nomothetical/ideographically) procedures to collect information. It is an intellectual task of creating concepts and ways of collecting and interpreting data in order to consider the complexities of the situations that we confront in teaching and teacher education; it involves theoretically attention to structural relations in which schooling exists; while at the same time giving reference to the historical specificity of our human conditions (see, e.g., Mills, 1959; Wallenstein, 1991). Further, the concepts of evaluation need to provide ways of considering the complexities of knowledge

While I have not exhausted any possible set of questions for considering issues of the Enlightenment in schooling, I recognize that the paradoxes of knowledge and power relations also produce paradoxes for evaluation. The imposition of a curriculum assumes a transcendence of certain knowledge that has a potential for achieving a better society; yet to propose a single form of knowledge is to structure out other possibilities. This process is never neutral and never without social implications. Knowledge is always located in a social and material world. The contradictions of teaching and teacher education are those of our occupational roles. In focusing on the issues of Enlightenment as frames for evaluation, we return to the problem of irony, contradiction, and dilemma that evaluations can illuminate.

V. Conclusion

In fundamental ways, teacher education is bound to tensions of violation, production, and reproduction in society. Schooling is a social creation to deal with the ruptures of cultural production and reproduction (see Lundgren, 1989). For many in our contemporary landscape, schooling is part of the modern quest to eliminate inequality and injustice; at the same time, there are the larger tensions of the structure of inequality that occurs in the cultural debates of school. While certain groups in the United States call for cultural pluralism as a way to give focus to the integrity of disenfranchised groups, others are calling for a new nation-building effort for U.S. schools. For the former, pedagogy is to make distinctions and difference as a valued category of society. The latter fears the increasing minority population in schools and suggests that schools strive to help create a national consensus and social solidarity. For these people, recognizing cultural differences is a tactic for arriving at more varied (and in the aggregate more effective) methods of putting across the traditional curriculum. With scores on standardized tests as the measure of success, schooling retains the particular cultural discourse that is embedded in the standardized testing industry. Teacher education, in this context, carries the tensions, violations, and productive elements of schooling itself.

We like to pretend that the world can be made rational, that progress is an obtainable goal, and that policy is the instrument of the modern version of salvation. I do not deny that we must keep on trying, but I also recognize that we know little about social and educational change. A focus on ambiguity and uncertainty is my way to explore what Lundgren has called the tension between our hopes and happenings.

I have located some questions about the social transformations, systems of ordering, and constructions of teacher education as a central problematic of evaluation. I recognize that these problems are not easily measured or conceptualized. It is almost as if our role is like that of Sisyphus -- never fully succeeding, but struggling to give attention to what is most important.

Notes

1. This paper was prepared for the Second National Research Symposium on Limited English Proficient (LEP) Students' Issues, sponsored by the U.S. Office of Bilingual Education and Minority Language Affairs and the Center on Assessment, Evaluation, and Testing at the University of California and the Center for Research on cultural Diversity and Second language learning, at the University of California, Santa Cruz. Washington, D.C.; Grand Hyatt Hotel, September 4-6, 1991.

2. This is contrary to the argument of Berger and Luckman (1967) who separate primary and secondary institutions of socialization, defining school as the latter.

3. I recognize that assessment procedures tied to science particularly those of psychometry were created with the development of mass schooling in the United States. But the tying of reform and evaluation as a state strategy is institutionalized after World War II.

4. The impulse for reform is so powerful in the educational field that it is practically impossible to distinguish research from evaluation. The name of the current research "game" is to privilege what is thought to lead to improved school practice. Over 20 nationally funded research centers exist as part of the current effort toward school reform. A task of many of these centers is to search for exemplary schools and teacher education programs, for example, and to explicate their characteristics. The assumption is that qualities of "good" schooling can be identified and exported to other schools.

5. The work of the Umea University Group on Evaluation, led by Professor S. Franke-Wikberg, has called this approach "theoretically organized" evaluation (see, e.g., Franke-Wikberg, 1982, 1990).

6. In this project, I have worked with Sigurjon Myrdal, Jay Hammond Cradle, Seehwa Cho, and Jim Ladwig.

7. This finding is consistent with teacher socialization literature; see, e.g., Zeichner & Gore, 1990.

8. Charles Bruckerhoff (1990) discusses this phenomenon in urban settings.

9. IGE=Individually Guided Education; ICC=Instructional Coordinating Committee.

10. These theoretical concerns can be found in feminist theory, although focused upon a particular social arena. See Nicholson, 1986, and Weedon, 1987.

11. This and the following section are drawn from Popkewitz, 1990.

References

Ball, D. (1988). Research on teaching mathematics: Making subject matter knowledge part of the equation (Research Report 88-2). East Lansing, MI: The National Center for Research on Teacher Education.

Berger, P., Berger, B., & Kellner, H. (1973). The homeless mind: Modernization and consciousness. New York: Vintage.

Billig, M. (1982). Ideology and social psychology. Extremism, moderation and contradiction. New York: St. Martin's Press.

Bourdieu, P. (1991). Language and symbolic power. (J. Thompson, Ed.), Cambridge, MA: Harvard University Press.

Bourdieu, P. (1989). Homo academicus. Stanford, CA: Stanford University Press.

Bourdieu, P. (1984). Distinction: A social critique of the judgment of taste (R. Nice, Trans.). Cambridge, MA: Harvard University Press.

Burrage, M., & Torstendahl, R. (Eds.). (1990). Professions in theory and history: Rethinking the study of the professions. Newbury Park, CA: Sage.

Cherryholmes, C. (1988). Power and criticism: Post-structural investigations in education. New York: Teachers College Press.

Clifford, J. (1988). The predicament of culture: Twentieth century ethnography, literature and art. Cambridge, MA: Harvard University Press.

Collins, R. (1990). Changing conceptions in the sociology of the professions. In R. Torstendahl & M. Burrage (Eds.), The formation of professions; knowledge, state and strategy. London: Sage Publications.

Dreyfus, H., & Rabinow, P. (1983). Michel Foucault: Beyond structuralism and hermeneutics, Chicago, IL: University of Chicago Press.

Foucault, M. (1988). The political technology of individuals. L. Martin, H. Gutman, & P. Huttan (eds.). In Technologies of the self: A seminar with Michel Foucault. Amherst: University of Massachusetts Press.

Franke-Wikberg, S. (1982). From ideas on quality of education to ideas as quality of education. Paper presented at a conference on the Socialization Effects of Higher Education, Klagenfurt, West Germany.

Franke-Wikberg, S. (1990). The American research on educational evaluation and the national evaluation in Sweden. In M.

Granheim, M. Kogan & U. Lundgren (Eds.). (1990). Evaluation as policymaking: Introducing evaluation into a national decentralized educational system (pp. 155-171). London: Jessica Kingsley.

Giddens, A. (1987). Social theory and modern sociology. Stanford, CA: Stanford University Press.

Haskell, T. (1984). Professionalism versus capitalism: R. H. Towney, Emile Durkheim, and C. S. Pierce on the disinterestness of professional communities. In T. Haskell (Ed.), The authority of experts: Studies in history and theory (pp. 180-225), Bloomington: Indiana University Press.

Holmes Group. (1986). Tomorrow's teachers. East Lansing, MI: Author.

Holmes Group. (1990). Tomorrow's schools. East Lansing, MI: The Holmes Group.

Kvale, S. (1990). Evaluation and decentralization of knowledge. In M. Granheim, M. Kogan & U. Lundgren (Eds.), Evaluation as policymaking: Introducing evaluation into a national decentralized educational system (pp. 119-140). London: Jessica Kingsley.

Labaree, D. (1990). Power, knowledge, and the science of teaching: A genealogy of teacher professionalization. Unpublished paper. Michigan State University, School of Education, East Lansing, MI.

Lundgren, U. (1990). Educational policy-making, decentralization and evaluation. In M. Granheim, M. Kogan & U. Lundgren (Eds.), Evaluation as policymaking: Introducing evaluation into a national decentralized educational system, London: Jessica Kingsley.

Lundgren, U. (1989). Educational research and the language of education. In D. Ball and S. Larsson (Eds.), The struggle for democratic education, equality, and participation in Sweden (pp. 191-210). New York: Falmer Press.

Lundgren, U. (1983), Between hope and happening: Text and context in curriculum. Geelong, Australia: Deakin University Press.

Mattingly, P. (1987). Workplace autonomy and the reforming of teacher education. In T. Popkewitz, (Ed.), Critical studies in teacher education: Its folklore, theory and practice (pp. 36-56). New York: Falmer.

Meyer, J. (1987). Self and Life Course: Institutionalization and Its Effects. In G. Thomas, J. Meyer, F. Ramirez, & J. Boli, (Eds.). Institutional Structure: Constituting State, Society, and the Individual (pp. 242-260). Newbury Park, CA: Sage Publications.

Miller, L. S. (1986, May 14). Nation-building and education, Education Week, p. 40.

Mills, C.W. (1959). The sociological imagination. New York: Oxford University Press.

National Council of Teachers of Mathematics. (1989), Curriculum and evaluation standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.

Noujain, E. (1987). History as genealogy: An exploration of Foucault's approach. In A. Griffiths, (Ed.), Contemporary French philosophy. (pp. 157-174). New York: Cambridge University Press.

O'Donnell, J. (1985). The origins of behaviorism: American psychology, 1876-1920. New York: University Press.

Popkewitz, T. S. (1991). A political sociology of educational reform: Power/knowledge in teaching, teacher education, and research. New York: Teachers College Press.

Popkewitz, T. (1984). Paradigm and ideology in educational research, social functions of the intellectual. London and New York: Falmer Press.

Popkewitz, T., & Pitman, A. (1986). The idea of progress and the legitimation of state agendas: American proposals for school reform. Curriculum and Teaching, 1(1-2), 11-24.

Popkewitz, T., & Lind, K. (1989). Teachers' incentives as reform implications for teachers' work and the changing control mechanism in education. Teachers College Record, 90(4), 575-594.

Popkewitz, T., Tabachnick, B., & Wehlage, G. (1982). The myth of educational reform: A study of school responses to a program of change. Madison: University Of Wisconsin Press.

Powell, A. (1980). The uncertain profession. Harvard and the search for educational authority. Cambridge, MA: Harvard University Press.

Rachjman, J. (1985). Michel Foucault: The freedom of philosophy. New York: Columbia University Press.

Ross, D. (1991). The origins of American social science. New York: Cambridge University Press.

Rorty, R. (1989). Contingency, Irony and Solidarity. New York: Cambridge University Press.

Schneider, B. (1987). Tracing the provenance of teacher education. In T. Popkewitz, (Ed.), Critical studies in teacher education: Its folklore, theory and practice (pp. 211-242). New York: Falmer.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. Harvard Educational Review, 37(1), 1-22.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4-14.

Smith, D. (1990). Conceptual practices of power: A feminist sociology of knowledge. Boston: Northeastern University Press.

Sykes, G. (1990). Evaluation in the context of national policy and local organization. In M. Granheim, M. Kogan & U. Lundgren (Eds.), Evaluation as policymaking: Introducing evaluation into a national decentralized educational system, London: Jessica Kingsley. (pp. 179-196)

Tom, A.R. & Valli, L. (1990). Professional knowledge for teachers. In W. Robert Houston (Ed.), Handbook of research on teacher education (pp. 373-392), New York: Macmillan.

Wallenstein, I. (1991). Unthinking social science: The limits of nineteenth-century paradigms. Cambridge, U.K.: Policy Press.

Weiler, H. N. (1990). Decentralization in educational governance: An exercise in contradiction? In M. Granheim, M. Kogan & U. Lundgren (Eds.), Evaluation as policymaking: Introducing evaluation into a national decentralized educational system, (pp. 42-65). London: Jessica Kingsley.

Zeichner, K. & Gore, J. (1990). Teacher socialization. In W. Robert Houston (Ed.), Handbook of research on teacher education (pp. 329-348), New York: Macmillan.

Assessing Appropriate and Inappropriate Referral Systems for LEP Special Education Students

Alba A. Ortiz
University of Texas, Austin

By the year 2,000, the United States will have 260,000,000 people, one of every three of whom will be African American, Hispanic, or Asian American. Minority students will comprise the majority of public school students, especially in large city schools. Students from minority groups already account for more than 50 percent of K-12 school enrollments in seven states (Individuals with Disabilities Education Act [IDEA] of 1990).

These demographic changes have focused attention on the educational status of multicultural populations. Unfortunately, the overwhelming evidence is that minority students experience limited academic success. For example, Gottfredson (1988) found that urban systems retain 15-20 percent of at-risk students at each grade level and that by the 10th grade, 60 percent of these students have been retained at least once. Retention is a common response to academic failure, even though there is little data to suggest that it leads to improved performance. On the contrary, data suggest that retention significantly increases the probability that students will dropout before graduation (Natriello, McDill, & Pallas, 1990). The dropout rates for minorities is 68 percent higher than for Anglo students (IDEA, 1990). A recent report of the National Commission on Secondary Schooling for Hispanics (1984) indicated that 45 percent of Mexican American and Puerto Rican students who enter school never finish and that of *all* Hispanics, 40 percent who leave school do so before tenth grade. Of Hispanics who took the "High School and Beyond" achievement tests, 76 percent scored in the bottom half of the national norms; it is not surprising, then, that 40 percent of the Hispanic student population is in a general education, versus an academic track.

Lack of educational progress of Hispanics and other language minority students has very important implications for special education as these students are likely to be referred for special services. More minorities continue to be served in special education than would be expected from their percentage of the general school population. Language minorities are overrepresented in programs for the learning disabled and, with the exception of Asian students, underrepresented in programs for the gifted and talented. With projections that one of every three Americans in this country will be black, brown, or Asian by the year 2,000, greater attention must be given to assuring that multicultural populations succeed in mainstream education and that procedures used to assess functioning levels and to recommend services reflect that those involved in the decision-making process understand how language and culture influence performance.

Otherwise, the increasing diversity of students in today's schools will overwhelm special education programs (Phillips and McCullough, 1990).

Issues Associated with Referral of Students to Special Education

Algozzine, Christenson and Ysseldyke (1982) conducted a national survey of directors of special education and asked them how many students had been referred between 1977 and 1980. The authors found that from 3 to 6 percent of the school-age population was referred each year for assessment. Of those referred, 92 percent were tested and 73 percent were found to be eligible for special education services. Ysseldyke, Thurlow, Graden, Wesson, Algozzine, and Deno (1983) conclude:

> It is clear that the most important decision made in the entire assessment process is the decision by a regular classroom teacher to refer a student for assessment. Once a student is referred, there is a high probability that the student will be assessed and placed in special education (p. 80).

While some would argue that there is no harm in placing students in special education who are already failing in the regular classroom, Wilkinson and Ortiz (1986) found that, after three years of special education placement, Hispanic students who were classified as learning disabled had actually lost ground. Their verbal and performance IQ scores were lower than they had been at initial entry into special education and their achievement scores were at essentially the same level as at entry. Neither regular education nor special education programs adequately served the academic needs of these language minority students.

An issue more basic than whether students profit from special education placement is whether they are eligible for such services in the first place. Algozzine and Ysseldyke (1981) found that 51 percent of placement team decision makers declared *normal* students eligible for special education services. Shepard (1987) and her colleagues (Shephard, Smith, & Vojir, 1983) estimate that half of the learning disabled population can be more accurately described as slow learners, second language acquirers, naughty children, students who are absent and move from school to school or average children in above average school districts. Shepard and Smith (1981) contend that half of the students placed under the label of perceptual and communication disorders (PCD) are misplaced:

> ...half of the children currently placed as PCD do not qualify by any definition of handicap. The most serious issue to be considered in response to this finding is that many of the "non-handicapped" children have serious problems in school and need special help. This is especially true for pupils in the language interference group....They may lag seriously behind in school because their first language is not English or because they may have trouble adapting to the mores of the school....They are not handicapped, yet they need extra attention, and there is currently no way to provide it other than labeling the child PCD (p. 170)

These data suggest that children with no readily identifiable handicapping condition are being considered for special education placement in increasing numbers. In fact, research shows that teacher referrals are often based on such extraneous factors as race, sex, physical appearance, and socioeconomic status as opposed to the pupil's need for special services (Bennet & Ragosta, 1984). In the case of limited English proficient (LEP) students in programs for the learning disabled (Cummins, 1984; Ortiz et al., 1985) and the speech and language handicapped (Ortiz, García, Wheeler, & Maldonado-Colón, 1986), neither the data gathered as part of the referral and evaluation process, nor the decisions made using these data, reflect that professionals adequately understand limited English proficiency, second language acquisition, cultural, and other differences which mediate students' learning.

In addition to evidence that the background characteristics of students influence referral, there is a growing body of literature indicating that many students served in special education experience difficulties which are "pedagogically induced" (Cummins, 1984). According to Hargis (cited in Gickling & Thompson, 1985):

> These children, who are in fact the curriculum casualties or curriculum handicapped, would not have acquired their various labels had the curriculum been adjusted to fit their individual needs, rather than having tried to force the children to achieve in the artificial but clerically simpler sequence of grades, calendar and materials that comprise the curricula. (p. 209)

Although there is often a requirement that the individual initiating the referral document interventions tried to improve academic performance prior to referral to special education, this is frequently not done (Gartner, 1986). Gartner concludes that we have the worst of alternatives in place: (a) a process that makes it easy to refer a student, with no check as to whether the referral may be a matter of prejudice against the child or failure on the school's part to meet the child's need, and (b) a system which not only does not demand but, in fact, provides little incentive for "prevention." He laments this situation because of his strong belief that most special education students could be better served in a general education system that gives greater attention to individual needs, adapts learning environments to accommodate diversity, provides training and support to increase the ability of school staff to respond to student diversity, and which funds efforts aimed at prevention rather than allocating resources to costly remedial programs.

The purpose of this paper is to discuss both referral and prereferral processes and to suggest how these might be made more effective. By design, more attention is given to prereferral intervention because available literature on the topic of special education referral consistently recommends that the best way to improve referral practices is to begin by implementing effective prereferral strategies. When regular educators, including bilingual education and English as a second language programs and personnel, respond to the unique needs of students, fewer of these students will need to be referred to special education. Those that are likely to be eligible for services because prereferral interventions will have exhausted all possibility that they can be maintained in the mainstream without specialized assistance.

Prereferral Intervention: Prevention

Prereferral intervention attempts to deal with learning and behavior problems that might otherwise be inaccurately identified as disabilities, at the site of their emergence -- the regular education classroom (Pugach & Johnson, 1988). In practice, prereferral intervention generally refers to a teacher's modification of instruction or classroom management, before referral, to better accommodate difficult-to-teach students who are not disabled (Fuchs, Fuchs, Bahr, Fernstrom, & Stecker, 1990). With increasing frequency, prereferral processes are also designed to minimize inappropriate referrals by strengthening the teacher's capacity to intervene with a greater diversity of student background characteristics, skills, abilities, and interests.

This traditional definition of prereferral intervention may be too narrow to adequately address the widespread failure of minority students in today's schools. The search for the cause of school-related difficulties should begin with an examination of whether students have been provided a school and classroom context conducive to success -- a context which reflects understanding and acceptance of linguistic and cultural diversity and other student characteristics and a curriculum appropriate to the needs of the learner, teachers, and other service providers who have direct training and experience in teaching multicultural populations. This suggests that prereferral intervention should be conceptualized as having two major components: (a) a prevention component aimed at establishing educational environments conducive to the academic success of language minority students so that problems will not occur in the first place, and (b) a problem-solving component in which the teacher first adapts instruction and/or the classroom environment to improve student performance and then requests assistance from others if problem-solving efforts are not successful.

A Framework for Empowering Minority Students. Prevention begins with establishing an educational environment that fosters success rather than breeds failure among minority students. Cummins (1986) argues that educational reforms which have attempted to reverse the pattern of underachievement and failure among minority students in the United States have been largely unsuccessful because they have not altered the historical relationships that have existed between teachers and students, and between schools and communities. To reverse the trend of widespread failure, educators, especially teachers, must redefine their roles within the classroom, the community, and the broader society so that these role definitions result in interactions that empower, rather than disable, students. Such redefinition is an important aspect of the first component of prereferral intervention -- preventing problems from occurring in the first place.

Cummins describes educators' role definitions along a continuum with one end promoting the empowerment of students and the other contributing to the disabling of students. Disabled students are considered as inherently inferior and are characterized by low achievement, high drop out rates, and high rates of referral to special education. In contrast, students who are empowered by their school experiences develop the ability to succeed. Cummins' framework for empowerment of minority students is summarized briefly below.

Collaborative school-community relationships. Schools are influenced greatly in their relationship with minority communities by the power and status relationships between minority and majority groups in the larger societal context (Fishman, 1976; Ogbu, 1978; Paulston, 1980). When societal conditions do not permit positive orientations between home and school, minority students come to school already predisposed to failure, a situation exacerbated by parents' limited access to economic and educational resources, bicultural ambivalence, and interactional styles that may not facilitate successful teacher-student interactions in the classroom (Heath, 1983; Wong-Fillmore, 1983).

Failure can be prevented if minority groups are positively oriented toward both their own and the mainstream culture, and if they do not perceive themselves to be inferior to the dominant group. Teachers with an exclusionary orientation tend to view parental involvement as either irrelevant or detrimental to children's progress. On the other hand, teachers who want to empower students, attempt to actively involve parents and other community members in the schooling process. Collaborative approaches between school and home allow parents to develop a sense of their own effectiveness in relation to their children's education, which, in turn, results in students' increased interest in school learning as well as improvement in behavior. To achieve an inclusive orientation, teachers must actively encourage parent involvement in their child's education both at home and at school. Moreover, if they are not bilingual, they must be willing to work closely with other teachers and aides who speak the child's primary language or dialect in order to communicate effectively.

Cultural and linguistic incorporation. Historically, "compensatory" education programs have been used by educators to equip minority students with academic and language skills required for success in mainstream society. However, by their very nature and orientation, compensatory programs are designed to *replace* minority students' primary language and culture with those skills deemed more critical to later social, economic, and academic success (e.g., the acquisition of English proficiency and knowledge of the dominant culture). When instruction is at the cost of the student's own culture and language, it is subtractive and defeats the very goals it seeks to accomplish.

In contrast to the subtractive orientation, additive approaches incorporate CLD students' culture and language in the teaching-learning process, communicate value and respect for the students' own diverse backgrounds, and reinforce their cultural identity, while at the same time teaching critical language, academic, and social skills. In schools that empower minority students, educators and the materials they use go beyond attempts to incorporate traditional aspects of the student's culture (e.g., food, music, festivals, and clothing) into the curriculum, since these aspects frequently fail to acknowledge the contemporary social, political, and economic experiences of minority groups. Moreover, such attempts are often characterized by fragmentation and isolation and may communicate stereotypes of racial and ethnic groups.

The curricula and instructional materials should be reviewed to determine whether they present both minority and majority perspectives and contributions and to determine whether they are relevant to students' language and culture. If student failure can be attributed to the use of inappropriate curricula or to ineffective instructional materials, then referrals to special education are unwarranted.

Efforts, instead, should focus on modifying or creating more effective instructional programs.

Instruction should be consistent with what is known about language acquisition and about the interrelationship between the first and the second language development (Garcia & Ortiz, 1988). Teachers should mediate instruction, using both the first and the second language, and integrate English development with subject matter instruction. Along with this, they should also respond to, and use, cultural referents during instruction, respecting the values and norms of the home culture even as the norms of the majority culture are being taught (Tikunoff, 1985).

The research literature (Cummins, 1984; Krashen, 1982) indicates that the native language provides the foundation for acquiring English as a second language skill. Therefore, educational programs which empower students have strong special language programs which promote native language conceptual skills as a basis for English communicative competence and literacy development, Cummins, 1984). Conversely, programs which prematurely shift students into English-only instruction interrupt a natural developmental sequence and interfere with intellectual and cognitive development. It is this interference that leads to academic failure and eventual referral to special education.

Interactive pedagogical approaches. Cummins believes that most curriculum planning in North America is characterized by a "transmission" model of instruction. Transmission-oriented teaching emphasizes sequential learning objectives, based on analysis of academic task demands, and directs instruction on these individual task components. Cummins argues that structuring learning into small, sequential steps tends to strip activities of the context required for that learning, thereby removing all cues that the child would need in the active generation of meaning. By structuring and grading learning experiences, the teacher becomes the initiator and controller of interactions with students, further stripping the learning situation of student control and intrinsic motivation. Teacher control assigns a passive role to the child, which further inhibits the intrinsic motivation and active involvement in learning that are essential for the development of higher order cognitive and academic skills. Thus, these models serve to maintain students' low functioning.

Cummins proposes, instead, that interactive approaches be used for instruction of language minorities. These approaches incorporate the basic tenets of language and literacy acquisition reflected in current research in these areas: (a) genuine dialogue between teacher and student in both oral and written modalities; (b) guidance and facilitation rather than control of student learning by the teacher; (c) encouragement of student-student talk in a collaborative learning context; (d) encouragement of meaningful language use rather than correctness of surface forms; (e) conscious integration of language use and development into all curricular content; (f) a focus on developing higher-level cognitive skills rather than basic skills; and (g) task presentations that foster intrinsic, rather than extrinsic, motivation.

Instruction should be consistent with what is known about language acquisition and about the interrelationship between the first and the second language development. The research literature (Cummins, 1984; Krashen, 1982) indicates that the native language provides the foundation for acquiring English as a second language skill. Therefore, strong promotion of native language conceptual skills will be more effective in providing a basis for English literacy (Cummins, 1984). Conversely, a premature shift to English-only instruction interrupts a natural developmental sequence and interferes with intellectual and cognitive development. Teachers should mediate instruction, using both the first and the second language, and integrate English development with subject matter instruction. Along with this, teachers should also respond to, and use, cultural referents during instruction, respecting the values and norms of the home culture even as the norms of the majority culture are being taught (Tikunoff, 1985). Above all, teachers must communicate high expectations for students and a sense of efficacy in terms of their own ability to teach culturally and linguistically diverse students.

Advocacy-oriented assessment. As indicated previously, a review of the referral-assessment-placement literature has also suggested that once a student in referred for special education, there is a high probability (75-90 percent) that he or she will be identified as handicapped (Reynolds, 1984). The assessment process has traditionally served to legitimate the disabling of minority students (Cummins, 1986). Because medical models are predisposed to locating psychological dysfunction within the student, ecological models of assessment are needed whereby the learning problem is examined in light of all contextual variables affecting the teaching-learning process, including teachers, students, curriculum, instructional approaches, and so forth. In the Cummins framework, an advocacy-oriented or "delegitimization" role for assessment personnel would involve "locating the pathology within the societal power relations between dominant and dominated groups, in the reflection of these power relations between school and communities, and in the mental and cultural disabling of minority students that takes place in classrooms" (Cummins, 1986, p.30).

Cummins' notion of advocacy-oriented assessment is compatible with the concept of prereferral intervention. In systems that empower students, teachers have the knowledge and skills to provide instruction consistent with students' needs. Moreover, they are adept at analyzing student performance, identifying gaps in skills and knowledge, and developing instruction to remediate those gaps within the framework of reciprocal interaction teaching. The importance of clinical teaching is discussed in a later section as it is an important component of prereferral intervention for students experiencing academic and behavioral problems in the regular education classroom.

Stedman's Formula for Effective Schools for Minority Students

Stedman argues that recent educational reforms which are based on the traditional effective schools' formula have resulted in a narrowing of the curriculum in a quest for higher test scores, neglect of higher-order thinking skills and liberal arts subjects, and increased teacher burnout. He cautions that implementation of the effective schools formula in low-income, urban schools may lead to a widening gap between the academic achievement of minority students and that of their Anglo peers. Moreover, Stedman questions how traditional approaches to schooling, which have proven unsuccessful in the past, can now be expected to produce academic success for all students.

The effective schools' literature delineates a set of factors believed to correlate positively with student gains in achievement. These factors include strong leadership by the principal, high expectations for student achievement, emphasis on basic skills, an orderly environment, systematic evaluation of students, and increased time on task (Stedman, 1987). Stedman analyzed case studies of schools which achieved grade-level success with low-income students and which maintained this success over several years. Based on this analysis, he offers a new synthesis of the effective schools' literature and a more practical approach to school improvement. Stedman's formula parallels very closely those factors included by Cummins (1986) and provides a data base to support this theoretical framework for empowering minority students. The alternative formula includes nine broad-based categories of highly interrelated practices.

Like Cummins, Stedman suggests that effective schools value *cultural pluralism* and acknowledge the ethnic and racial identity of their students and reinforce this identity by providing role models, offering bilingual education, and orienting students and their families to the school context. Effective schools provide mechanisms for administrator-parent-teacher-student *collaboration in governance*, rather than relying solely on the principal for instructional leadership. School personnel communicate frequently with parents (for example, through newsletters and home visits), encourage *parental involvement* in their children's learning, and provide opportunities for parents to participate in school governance. Lower teacher-pupil ratios are achieved in large part because positive school-community relations increase the number of volunteers and community resources available to students and provide more opportunities for adult-student interaction. In this way, *extra attention* can be given to students experiencing academic difficulty.

Students are actively engaged in their own learning through *academically rich programs* and tasks that capitalize on their personal experiences. Teaching is neither narrow, standardized, nor drill-based; basic skills are attained without sacrificing higher-order cognitive skills or a liberal arts education. Students are given responsibilities for student affairs and are involved in school governance. Good discipline is the result of the schools' organization and of their *positive, culturally-inviting learning environments*. Effective schools are "happy places," provide encouragement to students, and are not accepting of teacher unkindness.

In effective schools, the *best teachers are assigned those positions considered to be the most important*, including teaching in the early primary grades and remedial programs and serving as curriculum specialists or trouble-shooters. *In-service training is tailored to fit the specific needs of teachers and provides opportunities for them to share practical teaching techniques*. This fosters a collaborative learning community on the school campus.

Finally, *effective schools design their programs to ensure academic success and to head off academic problems*. For example, effective schools assign their best teachers to the early grades, sponsor home learning programs, lower the adult-pupil ratio, provide personal attention to students, and alert parents to their children's minor academic difficulties before they become serious problems.

Cummins and Stedman both suggest that the lack of success of educational reforms, especially those aimed at improving the education of minority students, may be due to the barriers that exist between educators and minority students and between schools and minority communities. Clearly, the message they communicate is that educational reform, in and of itself, is not sufficient for improving the educational status of minority students. Educators must create an educational context that is conducive to success and that communicates to students that they are valuable, competent individuals who can succeed in academic arenas. To provide an environment conducive to learning, school districts must endorse a philosophy of cultural pluralism and multicultural education, and instruction must reflect an understanding of how students' linguistic, cultural, and other background characteristics influence learning. Figure 1 provides an informal checklist which can be used to assess whether schools have been successful in providing this positive school climate which empowers minority students (Ortiz, 1988).

Figure 1
Evaluating the Educational Context

For each of the items below, circle "Yes" if the statement is characteristic of your school (or your district, if you prefer). circle "No" if the' statement is not characteristic of your school or district.

Yes No 1. My school/district supports cultural pluralism.

Yes No 2. The curriculum incorporates students' contemporary culture, not only history, customs and holidays.

Yes No 3. The curriculum helps students strike a balance between cultural pride and identity on one hand and appreciation of cultures different from their own on the other.

Yes No 4. The curriculum teaches certain humanistic values such as the negative effects of prejudice and discrimination.

Yes No 5. My school/district is integrated (or facilitates opportunities for cross-cultural interaction).

Yes No 6. Inservices routinely incorporate considerations in teaching linguistically/culturally diverse students.

Yes No 7. Children are encouraged to use their native language.

Yes No 8. The administration supports bilingual education.

Yes No 9. Minority parents are actively encouraged to participate in school activities .

Yes No 10. Training is provided to facilitate involvement of minority parents in their children's education.

Yes No 11. Parents and community members are given opportunities to provide input regarding important decisions.

Yes No 12. Parents and teachers participate in evaluations of school programs.

Yes No 13. Parents are considered to be valuable resources and are involved in the schooling process (e.g., as volunteers, advisory committee members, etc.).

Yes No 14. Standardized tests are used for special education eligibility decisions only if they are normed for multicultural populations.

Yes No 15. Regular classroom (not only bilingual education or ESL) teachers understand how limited English proficient students acquire English competence and incorporate language development activities in subject matter instruction.

Yes No 16. Minority students do as well on achievement tests as do Anglo students.

Yes No 17. Poor students do as well as middle- and upper-income students on tests of academic achievement.

Yes No 18. As much emphasis is given to developing higher cognitive skills as given to basic skill attainment.

Yes No 19. Teachers are facilitators of learning as opposed to transmitters of information and facts.

Yes No 20. Teachers adjust instructional approaches and activities to accommodate culturally-conditioned learning styles.

Yes No 21. Informal assessment is given as much emphasis as is formal assessment in psychoeducational evaluations of linguistically/culturally different students.

BEST COPY AVAILABLE

Yes No 22. Teachers are trained in informal assessment procedures.

Yes No 23. Reading and writing instruction is characterized by student control and an emphasis on meaningful communication and creativity.

Yes No 24. Teachers participate in decision-making.

Yes No 25. There is a well-articulated prereferral process in place to assure that students receive appropriate educational opportunities before they are referred to special education.

Yes No 26. The emphasis of assessment is on gleaning information to guide intervention.

Yes No 27. Students participate in school governance.

Yes No 28. Teachers are involved in planning and selecting inservice training topics and activities.

Yes No 29. Teams of educators, parents and community members participate in school improvement plans.

Yes No 30. My school/district would be described as a "happy" place by teachers and by minority students.

Prereferral Intervention: Problem Solving Processes

If students were to be provided positive school and classroom contexts that accommodate their individual differences or learning styles, most learning problems could be prevented. However, it is to be expected that even in these contexts, some students will experience difficulty. In these instances, teachers should cycle through a clinical teaching process in which they try several alternatives to resolve academic and behavior problems, including varying the instructional strategies and/or ensuring that the student has the necessary prerequisites to successfully complete tasks or assignments. If the teacher is unable to resolve the problem, she or he may need the assistance or support of others. If this is the case, it is important that teachers have access to a problem-solving process through which systematic efforts can be made to rule out all possibility that the student can be maintained in the regular classroom program.

Clinical Teaching

Before referring a student, teachers should carefully document adaptations of instruction and programs which have been attempted to improve performance in the mainstream (Garcia & Ortiz, 1988). Adelman (1970) suggests that instruction be carefully sequenced as follows: (a) teach basic skills, subjects, or concepts; (b) reteach skills or content using significantly different strategies or approaches for the benefit of students who fail to meet expected performance levels after initial instruction, and (c) refocus instruction on the teaching of prerequisite skills for students who continue to experience difficulty even after approaches and materials have been modified. Documentation of this teaching sequence is very helpful if the child fails to make adequate progress and is subsequently referred to special education. Referral committees will be able to judge whether the adaptations attempted were appropriate given the student's background characteristics. Ultimately, if the child qualifies for special education services, information about prior instruction is invaluable to the development of individualized educational programs because the types of interventions which work, and those which have met with limited success, are already clearly delineated.

When clinical teaching is unsuccessful, teachers should have immediate access to problem-solving units (Chalfant, Psych, & Moultrie, 1979). Otherwise, the simple passage of time may cause a problem to become so serious that it requires a special education referral. The most common problem-solving processes used by schools involve the use of consultants and/or problem solving teams for prereferral intervention.

Consultation Models

The consultation approach is meant to provide far more immediate service to classroom teachers in a far less structured manner than that involved in the use of problem-solving teams. There are two basic types of consultative models, expert and collegial; these are distinguished primarily by the level of shared knowledge or experience that initially exists among participants in the consultative process (Phillips & McCullough, 1990). In expert models, the relationship is hierarchical, with the consultant serving as the expert and the consultee receiving the expertise. In contrast, in a collegial model, peers join in exchanging specific ideas and experiences to solve problems encountered in areas of mutual understanding or interest.

In expert models, consultants typically offer the teacher advice as to how a problem may be resolved, provide direct intervention with the student, and/or guide him/her through problem identification, analysis, plan implementation, and problem evaluation (Fuchs, Fuchs, Bahr, Fernstrom, & Stecker, 1990). The consultant guides the teacher through these stages in a succession of structured interviews, in which specific objectives are accomplished before consultation proceeds to the next stage. Evaluation of interventions is data-based; effectiveness is judged in terms of whether the teacher has reached a previously set goal (e.g., changing the nature or quality of his/her interaction with students) or if the student's behavior has changed in the expected direction. In collegial relationships, the teacher is an equal participant in the process from problem identification to problem evaluation.

While the literature seems to favor collegial approaches to consultation, Fuchs, Fuchs, Bahr, Fernstrom and Stecker (1990) found that teachers prefer expert processes. In the first year of a study of prereferral intervention, they provided extensive training on collaborative consultation but lamented that the resulting in-class interventions were largely unimpressive. Teachers complained that they did not have adequate time to engage in the give-and-take nature of collaborative problem solving and simply wanted to be given helpful suggestions. When more prescriptive approaches were involved (that is, teachers were asked to select from among a limited set of carefully detailed interventions and development of prescriptive instructions and materials to guide them), teachers expressed satisfaction with the consultation process; they did not perceive the expert process to be coercive or denigrating.

Fuchs and his colleagues (1990) conclude that the form and substance of consultation should be consistent with the specifics of the situation. In schools where stress is high and expertise in consultation is not readily available, prescriptive approaches seem to be more successful than collaborative ones. As teachers and others become more confident and experienced in the process, the prescriptive approaches may give way to more collaborative efforts.

One of the major advantages of the consultant model is that teachers do not have to defend their perceptions of the problem before a public gathering of professionals as is typical of problem solving teams and/or referral committees (Pugach & Johnson, 1989). A disadvantage of the approach, though, is that the process, most often, relies on specialists for problem solution, thus creating a situation in which it is easier for teachers to transfer ownership of the problem to individuals they perceive as having specialized skills and knowledge. This is likely to be the case if the consultant assumes responsibility for generating solutions for the problem and then implementing them versus training the teacher to implement the strategy (Pugach & Johnson, 1989).

Problem-Solving Teams

Problem-solving teams generally serve two purposes: (a) they provide immediate, informal assistance to teachers to solve mild learning and behavior problems in the classroom, and (b) they serve as a screening mechanism for determining which students should be referred for a comprehensive individual assessment. Several alternatives for prereferral problem solving have been developed. These include among others, Child Study Teams, Student Assistance Programs, and Teacher Assistance Teams (TAT; Chalfant & Pysh, 1981). Members of the support team meet with the teacher requesting assistance to discuss presenting problems, brainstorm possible solutions, and develop an action plan that is then implemented by the teacher with the support of team members. The team conducts follow-up meetings to evaluate the effectiveness of the interventions and to develop other instructional recommendations if necessary. In many cases, it is the support team which ultimately decides whether the student should be referred to special education.

The following section summarizes how Teacher Assistance Teams operate (Chalfant & Pysch, 1981; Chalfant, Pysh & Maultrie, 1979). Although the focus of this particular process is on the student, it is also possible that presenting problems may be related to teacher variables or to the characteristics of the classroom environment.

Once the members of the TAT are elected, a team coordinator is named. The coordinator is responsible for overseeing data collection, scheduling meetings, and maintaining records of team meetings. Procedures used require minimal paperwork.

Teacher request for assistance. The teacher identifies a student-related problem and submits a brief, written summary of the problem to the TAT coordinator. The summary includes a description of (a) the performance the teacher desires of the child; (b) the students' strengths and weaknesses; (c) interventions already attempted and the outcomes of these; and (d) other relevant background information, including any available assessment data.

Review of requests for assistance. The TAT coordinator reviews the referral and, if necessary, confers with the referring teacher to clarify data or to obtain additional information about the problem. The coordinator then disseminates copies of the referral to the members of the committee. Team members review the information, pinpoint problem areas, study the interrelationships among these areas, and develop their own recommendations prior to the TAT meeting. This step reduces the amount of time spent discussing the dimensions of the problem at the meeting.

Classroom visits. One of the team members visits the classroom and observes the child to gather additional insights into the problem. While this step of Chalfant et al.'s process is child-centered, the committee should use this opportunity to gather information about the general classroom environment, including teachers, curriculum, and instruction.

Problem-solving meeting. A TAT meeting is held for 30 minutes at which time team members: (a) reach consensus as to the nature of the problem; (b) negotiate one or two objectives with the referring teacher; (c) select the methods, strategies, or approaches the referring teacher will attempt, (d) define responsibility for carrying out the recommendations (who, what, when, where, how, why); and (e) establish a follow-up plan to monitor progress.

Recommendations. The end products of the TAT meeting are specific recommendations for individualizing instruction for the student, recommendations for informal assessment to be conducted by the child's teacher or by team members, and/or referral for special help, including, if the team deems it necessary, referral to special education. Referrals for special help can be teacher-rather than child-focused. For example, an instructional strategy which is unfamiliar to the referring teacher may be recommended. The teacher can request in-service training to learn the strategy, other members of the faculty who have expertise in the recommended approach can demonstrate the strategy, or the team may recommend that the child be integrated into a classroom where such instruction is already being provided. The recommendations are recorded on a form during the meeting and xerox or carbon copies are provided to all team members.

Follow-up meetings. These meetings are held every six to eight weeks to review progress toward solving the problem. If the problem is resolved, techniques which can be used in similar cases are identified; if the interventions are not successful, the team repeats the brainstorming process and selects alternative strategies.

Referrals to other programs. If the LEP student's problems cannot be resolved by the bilingual education or ESL teacher, the TAT may refer the student to compensatory education programs which provide remedial instruction. Unless alternative placements such as these are readily available, referral to special education will continue to be a "trigger" response when teachers or problem-solving teams are unable to improve students' achievement or behavior. To access these alternatives, it is important that teachers understand their purpose and that they be familiar with eligibility criteria for placement (i.e., which students are served by which program). Otherwise, misplacement in special education can continue to occur despite the availability of options such as Chapter 1, migrant education, tutorial programs, and others. (García, 1984).

The quality of available programs must be carefully monitored as there is also well-documented overrepresentation of language minority students in programs such as Chapter 1. This does not usually cause concern or lead to litigation perhaps because these programs are assumed to be beneficial to students (Reschly, 1988) and do not carry the same stigma as a special education label or placement. However, overrepresentation suggests that the regular classroom environment is not effective for these students; rather than channeling students out of the mainstream, attention should be focused on improving these instructional environments.

There are several benefits to the use of Teacher Assistance Teams: teachers are provided a day-to-day peer problem-solving unit within their school building and thus do not have to experience long delays until external support can be provided (Chalfant, Psych, & Moultrie, 1979). Moreover, a collaborative learning community is established since the team process actually provides continual staff development for all persons involved in the process. Finally, the use of TAT serves to reduce the number of inappropriate referrals to special education because most problems can be taken care of by regular education personnel. An additional benefit of the TAT is that the process helps identify problem areas or training needs which, if addressed, can help school personnel deal more effectively with students' learning and behavior problems.

In summary, both consultation models and problem solving teams have been shown to be effective vehicles for operationalizing prereferral intervention. Educators are encouraged to explore the specific type of prereferral process which would be most effective given the characteristics of the school, its personnel, and available resources. It is important not to assume that only one combination of experts and their accompanying skills is adequate to address problems. Therefore, neither the consultation nor the team problem-solving model should be constituted as permanent structures (Graden, 1989).

Institutionalizing Prereferral Intervention

There are many benefits to be gained from the implementation of prereferral intervention strategies. The processes used for problem solving endorse the rights of teachers to assistance and support from colleagues and the educational system (Phillips & McCullough, 1990). They also underscore that such assistance should be provided in a timely manner and that teachers should not have to wait for results of testing before taking action (Pugach & Johnson, 1989). Moreover, numerous opportunities are provided for enhancement of teachers' abilities to respond to the growing diversity of the school population, abilities that are critical, given the nation's changing demography (Phillips & McCullough, 1990). Of utmost importance, given the dramatically increasing number of students identified as being "at risk," is that prereferral intervention is more cost-effective than are remedial programs for students who are not disabled. Special education involves substantially greater expenditures (e.g., for students with mental retardation 1.75 to 2.5 times) than the expenditures per student in regular education or from $2,000 to $4,000 annually (Reschly, 1988). While this is an expenditure that is appropriate when the student is truly disabled, expending this level of resources on non-disabled students can bankrupt the educational system.

Despite the theoretical support of the need for collaboration, true interdisciplinary collaboration is not routinely occurring. School personnel must thus develop ways to institutionalize this type of effort. Several factors are critical to achieve this end (Phillips & McCullough, 1990, pp. 291-295):

1. School districts must adopt a philosophy endorsing the concept of prereferral intervention, both the prevention and the problem-solving process, and enact policies and procedures consistent with collaborative problem solving. The system must communicate that teacher or student problem resolution merits expenditure of time, energy, and resources.

2. Problem solving teams must develop a collaborative ethic. Central tenets of this ethic are joint responsibility for problems, joint accountability, and a belief that linking talents and resources is mutually advantageous to regular and to special education.

3. The understanding and support of administrators is crucial if prereferral intervention is to be institutionalized. Principals can exert tremendous influence on program success through clear communication of program purpose, goals, and expectations, promotion of a climate in which consultation is valued, provision of leadership and utilization of managerial strategies which facilitate program implementation and maintenance. Because informal prereferral structures are not effective, administrators must ensure that consultation and team meetings can occur routinely.

A number of conceptual and pragmatic barriers to consultation and team problem solving have been identified (Phillips & McCullough, 1990; Chalfant & Pysh, 1989; Moore, Fefield, Spira, & Scarlata, 1989). In order for collaborative consultation to occur, the historical separation of special education and regular education must be eliminated. For LEP students, greater collaboration and cooperation between bilingual education and "regular" regular education must be achieved, in addition to strengthening linkages with special education. Attitudinal barriers caused by the lack of understanding of the roles of programs and personnel must be eliminated. This presents a challenge to bilingual educators who have continuously struggled with having to explain the nature and purpose of special language programs not only to regular and special educators but also to the community at large.

In strengthening relationships across programs and personnel, it is important to recognize that the trend toward prereferral intervention may not be eagerly embraced by regular educators. Emphasis on mastery of content and skills which are then measured by standardized achievement testing may cause teachers to refer students to special education and other remedial programs as a way of improving the academic achievement and thus the test scores of the students in their classes. This suggests a need for meaningful involvement of staff at all levels in planning and decision making (Phillips & McCullough, 1990) in order to increase the likelihood of successful implementation of prereferral procedures.

Need for Additional Research on Prereferral Intervention

Prereferral intervention programs are increasing in popularity. In a recent survey of 49 state directors of special education, Carter and Sugai (1989) found that 23 state educational agencies required and 11 recommended prereferral intervention; 10 had no prereferral requirements. The broadening support for this type of intervention has occurred in the absence of an adequate data base to support its effectiveness (Fuchs, Fuchs, Bahr, Fernstrom, & Stecker, 1990). These authors conducted an ERIC search which produced only three empirical investigations; they knew of only eight additional published, pertinent studies.

Despite the limited number of studies, available results have been generally encouraging (Phillips & McCullough, 1990). Efficacy reviews of outcome research (e.g., Mannino & Shore, 1975; Medway, 1982; West & Idol, 1987) and meta-analyses of consultation studies (Medway & Updyke, 1985; Sibley, 1986) have revealed positive effects on attitudes and behaviors of consultants, consultees, and clients. Moreover, applied researchers (e.g., Graden, Casey, & Bonstrom, 1985; Ritter, 1978) have suggested that well-designed consultation programs may significantly reduce the number of referrals and the long-term need for consultation services. For example, there is evidence to suggest that prereferral interventions can resolve a significant proportion of behavioral and academic problems and thus eliminate the need for referrals to special education (Ortiz, 1990; Reschly, 1988; Chalfant, 1981; Chalfant, in press).

Chalfant & Pysh (1989) conducted a study of the outcome of 96 Teacher Assistance Teams. They found that of the 386 students staffed by the teams, only 82 or 21 percent were referred for special education services. Of these, 76 percent were found to be eligible. Teachers involved in the process rated the group process as very effective for problem solving and indicated interventions implemented resulted in improvement of student behavior and achievement. Teachers also lauded the moral support provided by their peers. Graden, Casey, and Bonstrom (cited in Carter and Sugai, 1989) conducted a study which showed that in four of the six participating schools, testing and placement rates were decreased significantly as a result of prereferral intervention and that teachers and principals perceived the process to be helpful to students. Ortiz (1990) investigated the use of problem-solving teams on four elementary school campuses in which the majority of students enrolled were Hispanic. She found that of 100 students staffed by these teams, 73 percent were helped without referral to special education. Reschly (1988) cautions that the degree to which results such as these are persistent within settings, maintained across time, and generalized across situations, remains to be established.

The Referral Process

The previous discussion of prereferral intervention is not intended to suggest that referrals to special education are never appropriate. If neither the teacher's adaptations nor the recommendations of consultants or problem-solving teams are effective, then referral to special education should be considered. The data collected through prereferral intervention becomes invaluable to special educators as they move to a comprehensive individual assessment and try to determine whether the student is handicapped and to diagnose the specific disability. The evidence most critical to determining eligibility will accompany the referral, i.e., verification that: (a) the school's curriculum is appropriate; (b) the child's problems are documented across settings and personnel not only in school but also at home; (c) difficulties are present both in the native language and in English; (d) the child has been taught but has not made satisfactory progress; (e) the teacher has the qualifications and experience to effectively teach the student; and (f) instruction has been continuous, appropriately sequenced, and has included teaching of skills prerequisite to success. A child who does not learn after this type of systematic, quality intervention is a likely candidate for special education. If the student is handicapped, the records maintained by teachers and team members can guide the development of the individualized education plan (IEP) as effective and ineffective strategies have already been identified.

Accessing Special Education Services

Referrals to special education indicate that a decision has been reached that the child cannot be served by regular education programs alone, and that she/he may have a disability. The referral process then represents an additional opportunity to determine whether the student's problems can be attributed to factors other than a disability.

Every district is required to have a process for screening referrals. In some instances, an individual will be given responsibility for screening; in others, a group of individuals serves as the screening committee. In either case, the information provided by the referral agent, who is usually the child's teacher, drives referral decisions.

Algozzine and Ysseldyke (1981) examined the extent to which decisions to classify a child as mentally retarded, learning disabled, or emotionally disturbed are influenced by data provided at the time of referral. Results indicated that although all students fell within the normal range, 51 percent of the decision makers in the study declared the students eligible for special education services. The authors conclude that decision makers place considerable weight on information provided in referral information and, as a result, fail to reject stereotypes engendered in the referral statement. The implication of this finding is critical for improving procedures associated with referral of language minority students. If the final placement decision is so heavily weighted by the original referring data, mainstream teachers who are unable to distinguish those students in their classrooms whose performance is indicative of normal second language development from those who exhibit a true handicapping condition risk making an inappropriate referral, thereby effectively resigning the referred student to special education placement.

Given this, referral information should help distinguish linguistic, cultural, and other student differences from disabilities. Referral data should include information such as the following:

1. The student's current educational status, including attendance, grades, achievement data, and classroom observations;

2. Results of the *home language survey*;

3. Up-to-date descriptions of the *student's use of the native language and English language, including measures of basic interpersonal communication skills and academic language proficiency* (Cummins, 1984);

4. Documentation of previous educational efforts and strategies provided for the student and the results of these efforts, including *participation in or consideration for other special programs* operated by the district;

5. Documentation of recent vision and hearing screening;

6. An updated general health history or documentation of recent medical evaluations;

7. *Other information reported or provided by parents.*

Documentation of any decisions made by the bilingual education placement committee should always be included for limited English proficient students.

Committee Process

Referral activities are conducted very much like the prereferral problem-solving process; a team is brought together to deliberate available information about a student and to render a decision as to whether the child should be referred for a comprehensive individual assessment. If the referral committee determines that the child is not eligible for special education services, they usually recommend alternatives such as the following: (a) adjusting the student's educational program, (b) returning the student to the regular classroom with teaching recommendations provided to support the teacher, or (c) referring the student for consideration by other programs. If it is concluded that a child is not eligible for special education services, the referral committee, or the placement committee if the decision is made after the comprehensive individual assessment, usually recommends that additional modifications of the child's educational program be made; and/or that the student be considered for placement in compensatory or remedial programs. If a prereferral process is in place, referral committees can access this support system so that the teacher can be given assistance with students who have educational needs but who cannot be served in special education.

Representation on Prereferral and Referral Committees

A major debate associated with the prereferral intervention is whether consultants and/or members of problem-solving teams should be regular or special educators. Chalfant and Pysh (1981) argue that Teacher Assistance Teams *should not* involve special education personnel (e.g., special education teachers or psychologists) or other specialists, except when they are invited to serve as consultants to the committee. The presence of principals and special educators on teams may create conflicts for teachers; for example, they may be threatened because the principal normally serves an evaluative role and teachers may worry that their request for assistance will be interpreted as lack of competence. They may interpret the presence of a special educator as indicating that a referral is imminent (Phillips & McCullough, 1990). As a matter of fact, Graden (1989) suggests that rather than ***pre***referral intervention, the problem-solving process should be called *intervention assistance*. She cautions that teachers may interpret the term prereferral as simply signaling a step or action that has to be taken before the actual referral is made, rather than as a process aimed at preventing unnecessary referrals from occurring.

While reliance on specialists is a common criticism of the use of consultation models for prereferral intervention, Graden (1989) takes issue with the description of special educators as experts who intimidate teachers and who are unable to collaborate by virtue of their roles and titles. She suggests that such a posture is counterproductive to establishing more effective linkages between regular and special education. Rather than categorizing individuals on the basis of their roles, greater attention should be given to the skills and background they have to offer.

Ortiz (1990) concurs but argues that while availability of peer support is more important that team membership, the success of teams comprised of regular classroom teachers suggests that greater consideration should be given to the use of such teams of regular classroom teachers for prereferral intervention. A committee structure in which membership involves only regular classroom teachers (at least a majority of) emphasizes that prereferral intervention is under the authority, and is the responsibility, of the ***regular*** education system. It is this authority which distinguishes the ***prereferral*** from the ***referral*** process. Moreover, relying on regular educators allows specialists to spend more time on tasks for which they are uniquely trained (e.g., conducting assessment, serving on special education referral committees, providing direct services to students with disabilities, etc.).

While it is argued that special educators and certain other specialists (e.g., principals) should not serve as consultants or be members of problem-solving teams, the prereferral process should involve individuals with expertise associated with the education of limited English proficient students. Such expertise will be very helpful as team members attempt to rule out any possibility that a student's problems might be the result of differences in language, culture, socioeconomic status, or to not having had opportunities to learn. While initially these individuals may be seen as having specialized knowledge and skills relative to second language acquirers, this expertise must be shared by all regular educators. Otherwise, it will be impossible to achieve the type of educational context described previously in which all personnel understand the influence of language, culture, and other background characteristics in order to prevent academic problems from occurring. Both the consultation and problem-solving team process are excellent vehicles for sharing this expertise and moving it to becoming commonplace knowledge on school campuses.

Individuals with bilingual education and English as a second language expertise should also serve on referral committees. Ideally, this representative would be the classroom teacher if the child is limited English proficient and/or a representative from the bilingual education placement committee. These individuals can help interpret student behavior in light of linguistic and cultural characteristics. They would also be of great assistance in obtaining information about the child from parents and in helping understand the prereferral-referral-assessment-placement process. Second language program personnel can also offer invaluable assistance in configuring an assessment process that will ensure that performance is the native and the English language is accurately described and to ensure that assessments provide data appropriate to programming for instruction in both the first and the second language.

Summary

The anticipated outcomes of the implementation of prereferral intervention strategies include: (a) a reduction in the number of students perceived to be "at risk" by regular classroom teachers because of teachers' increased abilities to handle the naturally occurring diversity of skills and characteristics of students in their classes; (b) reduction in the number of students referred to special education; (c) reduction in the number of students inappropriately labeled as handicapped, particularly in programs for the learning disabled; and (d) improved student outcomes, especially in oral language and literacy skills.

Serving students in the mainstream is more cost-effective than placing them in special education, especially if the student is not handicapped. More important perhaps are the long-term benefits to students themselves. They will have a greater chance of achieving their social, political, and economic potential because they are provided an appropriate education and are spared the stigma of an inaccurate special education label.

References

Adelman, H. (1970). An interactive view of causality. Academic Therapy, 6, 43-52.

Algozzine, B., Christenson, S., & Ysseldyke, J. (1982). Probabilities associated with the referral to placement process. Teacher Education and Special Education, 5(3), 19-23.

Asher, J. (1979). Learning another language through actions: The complete teacher's guidebook. Los Gatos, CA: Skyoak Productions.

Bennett, C. I. (1986). Comprehensive multicultural education: Theory and practice. Boston: Allyn and Bacon, Inc.

Bergen, J., & Smith, J. O. (1966). Effects of socioeconomic status and sex on prospective teachers' judgements. Mental Retardation, 4, 13-15.

Berliner, D. C. (1984). The half-full glass: A review of research on teaching. In P.L. Hosford (Ed.), Using what we know about teaching (pp. 511-77). Alexandria, VA: Association for Supervision and Curriculum Development.

Brown, G. H., Rosen, N. L., Hill, S. T., & Olivas M. A. (1980). The condition of education for Hispanic Americans. Washington, D.C.: National Center for Educational Statistics.

Carter, J., & Sugai, G. (1989). Survey on prereferral practices: Responses from state departments of education. Exceptional Children, d55(4), 1-3.

Cazden, C. (1984). Effective instructional practices in bilingual education. Washington, D. C.: National Institute of Education.

Chalfant, J. C. & Van Dusen Pysch, M. (1989). Teacher Assistance Teams: Five descriptive studies on 96 teams. Remedial and Special Education.

Chalfant, J. C. & Van Dusen Pysch, M. (1981, November). Teacher Assistance Teams--A Model for within-building problem solving, Counterpoint, 16-21.

Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy. Clevedon, Avon, England: Multilingual Matters.

Fuchs, D., Fuchs, L. S., Bauhr, M. W., Fernstrom, P., & Stecker, P. M. (1990). Prereferral intervention: A prescriptive approach. Exceptional Children, 56(6), 493-513.

García, S. B. (1984). Effects of student characteristics, school programs and organization on decision-making for the placement of Hispanic students in classes for the learning disabled. Unpublished doctoral dissertation, The University of Texas at Austin.

Gartner, A. (1986). Disabling help: Special education at the crossroads. Exceptional Children, 53(1), 72-76.

Gay, G. (1981). Interactions in the culturally pluralistic classroom. In J. A. Banks (Ed.), Education in the 80's: Multiethnic education. Washington, DC: National Education Association.

Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. Exceptional Children, 52, 205-218.

Gollnick, D. M., & Chinn, P. C. (1986). Multicultural education in a pluralistic society (2nd ed.). Columbus, OH: Merrill Publishing Company.

Good, T. L. & Brophy, J. E. (1973). Looking in classrooms. New York: Harper & Row.

Gottfredson, D.G. (1988). Business community/school partnerships. Baltimore, MD: Center for Research on Elementary and Middle Schools.

Graden, J. (1989). Redefining "prereferral" intervention as intervention assistance: Collaboration between general and special education. Exceptional Children, 56(3), 227-231.

Graden, J. L., Casey, A., & Christenson, S. L. (1985). Implementing a prereferral intervention system: Part I. The model. Exceptional Children, 51, 377-387.

Graves, D. (1983). Writing: Teachers and children at work. Exeter, NH: Heinemann Books.

Harth, R. (1982). The Feuerstein perspective on the modification of cognitive performance. Focus on Exceptional Children, 15(3), 1-12.

Heath, S. B. (1986). Sociocultural contexts of language development. In California State Department of Education, Beyond language: Social and cultural factors in schooling language minority students (pp. 143-186). Los Angeles, CA: Evaluation, Dissemination and Assessment Center, California State University.

Heller, K. A., Holtzman, W. H., & Messick, S., (Eds.). (1982). Placing children in special education: A strategy for equity. Washington, D. C.: National Academy Press.

Jackson, G. & Cosca, C. (1974). The inequality of educational opportunity in the Southwest: An observational study of ethnically mixed classrooms. American Educational Research Journal, 11, 219-229.

Kaskowitz, D. H. (1977). Validation of state counts of handicapped children. Volume II - Estimation of the number of handicapped children in each state. (DHEW, Bureau of Education for the Handicapped Contract No. 300-76-60513). Menlo Park, CA: Stanford Research Institute.

Krashen, S. (1982). Principles and practice in second language acquisition. New York: Pergammon Press.

Mattes, L., & Omark, D. (1984). Speech and language assessment for the bilingual handicapped. San Diego, CA: College-Hill Press.

Northcut, M. S., & Watson, D. (1986). Sheltered English teaching handbook. Carlsbad, CA: Northcutt, Watson, Gonzales.

Ortiz, A. A. (1984). Choosing the language of instruction for exceptional bilingual children. Teaching Exceptional Children, 16, 208-212.

Ortiz, A. A., García, S. B., Holtzman, W. H., Jr., Polyzoi, E., Snell, W. E., Jr., Wilkinson, C. Y., & Willig, A. C. (1985). Characteristics of limited English proficient Hispanic students in programs for the learning disabled: Implications for policy, practice, and research. Austin, TX: The University of Texas, Handicapped Minority Research Institute on Language Proficiency.

Ortiz, A. A., García, S. B., Wheeler, D., & Maldonado-Colón, E. (1986). Characteristics of limited English proficient students served in programs for the speech and language handicapped: Implications for policy, practice, and research. Austin, TX: The University of Texas, Handicapped Minority Research Institute on Language Proficiency.

Ortiz, A. A. & Maldonado-Colón, E. (1986). Recognizing learning disabilities in bilingual children: How to lessen inappropriate referrals of language minority students to special education. Journal of Reading, Writing, and Learning Disabilities International, 1(1), 47-56.

Ortiz, A. A., & Yates, J. R. (1983). Incidence among Hispanic exceptionals: Implications for manpower planning. Journal of the National Association for Bilingual Education, 7(3), 41-53.

Ortiz, A. A. & Yates, J. R. (1984). Linguistically and culturally diverse handicapped students. In R. S. Podemski, B. M. Price, T. E. C. Smith, G. E. Marsh II (Eds.), Comprehensive administration of special education (pp. 114-141). Rockville, MD: Aspen Publication.

Phillips, V., & McCullough, L. (1990). Consultation-based programming: Instituting the collaborative ethic in schools. Exceptional Children, 56(4), 291-304.

Ramírez, M. & Castaneda, A. (1974). Cultural democracy, bicognitive development, and education. New York: Academic Press.

Reschly, D.J. (1988). Minority MMR overrepresentation and special education reform. Exceptional Children, 54(4), 316-323.

Reynolds, M. (1984). Classification of students with handicaps. In E. W. Gordon (Ed.), Review of research in education (Vol. 11, pp. 63-92). Washington, DC: American Educational Research Association.

Rist, R. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. Harvard Educational Review, 40, 411-450.

Rosenshine, B. V. (1983). Teaching functions in instructional programs. Elementary School Journal, 83, 335-352.

Sadker, M. & Sadker, D., (1982). Between teacher and student: Overcoming sex bias in classroom interaction. In M. Sadker and D. Sadker. Sex equity handbook for schools. NY: Longman.

Saville-Troike, M. (1978). A guide to culture in the classroom. Rosslyn, VA: National Clearinghouse for Bilingual Education.

Shepard, L. (1987). The new push for excellence: Widening the schism between regular and special education. Exceptional Children, 53(4), 327-329.

Staton, J. (1987, April). New research on dialogue journals. Dialogue, 4(1), 1-24.

Terrell, T. (1983). The natural approach to language teaching: An update. In J. W. Oller, Jr., & P. A. Richard-Amato (Eds.), Methods that work: A smorgasboard of ideas for language teachers. Rawley, MA: Newbury House Publishers.

Tikunoff, W. J. (1985). Applying significant bilingual instructional features in the classroom. Rosslyn, VA: National Clearinghouse for Bilingual Education.

Tindal, G., Shinn, M.R., Rodden-Nord, K. (1990). Contextually based school consultation: Influential variables. Exceptional Children, 56(4), 324-336.

Tucker, J. (1981). Sequential stages of the appraisal process: A training manual. Minneapolis, MN: National School Psychology Inservice Training Network.

Ysseldyke, J., Algozzine, B,. Richey, L., & Graden, J. (1982). Declaring students eligible for learning disability services: Why bother with the data? Learning Disability Quarterly, 5, 37-44.

Zigmond, N., & Miller, S. (1986). Assessment for instructional planning. Exceptional Children, 52, 501-509.

Response to Alba Ortiz's Presentation

Ann C. Willig
Florida Atlantic University

In addressing issues related to the assessment of systems for referring language-minority (LM) children to special education, Dr. Ortiz mentioned four points which emphasize that schools in this country are failing to meet the needs of LM children: (1) there continues to be a very high dropout rate from school for LM children; (2) over-representation of LM children in special education continues, especially in the area of learning disabilities and communication disorders; (3) there are large numbers of LM children in special education classes who really don't belong there; and (4) many LM children assigned to special education show a progressive decline of scores on intellectual and achievement tests over their years in special education. These points highlight three major areas of need that must be resolved if special education is to provide LM students with appropriate services: first, the need to reduce inappropriate referrals to special education, second, the need to reduce inappropriate placements for those students who are referred, and third, the need for appropriate instruction in special education classrooms for LM children who truly need special education.

Reduction of Inappropriate Referrals

As Dr. Alba Ortiz implies, the root of the problem of over-referral to special education is not in the referral and assessment systems per se, but in the nature of the regular, or non-special education programs that are offered to LM students in this country. Behaviors and characteristics of LM students that precipitate their referral are frequently produced by inappropriate educational programs and instruction that does not meet their needs. Reduction of inappropriate referrals to special education will best be accomplished through the assessment and improvement of general education programs offered to all LM students -- a task that is outside the realm of special education.

A first step in the assessment of general education is the examination of schools for all the characteristics that Dr. Ortiz listed in her paper, that is, the promotion of collaboration with parents and communities, the provision of culturally relevant education using techniques of effective multicultural education, the building upon language and knowledge that children bring to school, and the provision of meaningful and comprehensible instruction. Inappropriate referrals to special education will continue until schools can receive acceptable grades on these characteristics of their programs for English-learning students.

There is recent information and research evidence related to two of the areas of effective schooling for LM students that I would like to comment upon. The first concerns language-learning needs of children and instructional methods, the second addresses the language of instruction as a mediating factor in parent involvement.

Language Needs and Language Instruction

Findings from one recent study have classroom implications that are best understood when one considers the language development process in children. Current theories of language and cognition suggest that these are developed through a process of trying to make sense of our environment and to figure out the rules that govern our world and lives. In the case of language, this process includes three basic steps. The first is listening to language in our environments and trying to sort out what we hear until it makes sense to us with some degree of consistency. This is similar to a scientist who conducts preliminary observations of a particular phenomenon and then tries to make sense of these observations. In trying to make sense of observations, one begins to form hypotheses about relationships between various phenomenon and to figure out rules that may govern the patterns observed. Children do this when learning language, whether it be their first, second, or third language, as is evidenced in part by grammatical over-regularization (Dale, 1976) in young children who say such things as "I doed it," "I didn't spilled it," and so forth. Children try to figure out the rules.

Following preliminary observations and the development of hypotheses, more systematic observations are done and hypotheses are tested out. These must be tested, not just once but many times, until repeated observations confirm the consistency of results, just as repeated experiments are performed in scientific disciplines to confirm consistency of results. Anyone who has listened to young children ask the same questions over and over or heard them repeat the same sounds or phrases in many different situations has observed this hypothesis testing.

In short, the process includes two stages of observation plus hypothesis-formation and hypothesis testing. When children learn a language, they must be provided with data that serve as a basis for observation and hypothesis formation, that is, they must hear meaningful language in their environment. Just as important, they must be provided with the opportunity to test their language-learning hypotheses through language production and interaction with other individuals. Language learning hypotheses are confirmed or disconfirmed by the way other individuals respond to these language production efforts.

The point of the above is that a crucial ingredient of appropriate instruction for English-learning students is the provision of opportunities to practice language and to test out language-learning hypotheses through real interaction with teachers and peers.

If we were to assess our schools on a grand scale for just this one ingredient of appropriate instruction for English-learning students, we would come up very short. Classroom evidence from several nationwide studies indicates that teachers provide little opportunity for students to produce language in meaningful ways.

The recent nationwide study conducted by David Ramirez and his colleagues (Ramirez, J.D., Pasta, D.J., Yuen, S.D., Billings, D.K., and Ramey, D.R., 1991) provides evidence in this respect. Ramirez and his colleagues set out to compare three types of programs for English-learning students -- immersion, early-exit, and late-exit bilingual programs.

In order to compare the effectiveness of these programs, the researchers had to examine the quality of instruction in each program to make sure that any program effects could be attributed to the program models and not to differences in the quality of instruction provided by the teachers. The result of classroom observations and careful documentation of teacher-student interactions indicated that, in each of the three types of program, approximately 95 percent or more of the classroom interactions were teacher-initiated and consisted of display questions, that is, questions that require responses of only a few words and that merely display memorization or rote recall of facts. Crucial language-production opportunities were not provided for students in these programs.

The finding itself is not surprising. What is surprising is that these observations were in programs especially designed to meet the needs of English-learning students. Failing to provide opportunities for meaningful language production means failing to meet the needs of these students.

That this finding is not surprising stems from other research which indicates that the above describes the prevalent mode of teaching in our country. Ventriglia (personal communication) described data from two nationwide studies in which she analyzed more than 13 million teacher and student interaction chains collected in regular education classrooms of standard English, ESL classrooms, and native language instruction classrooms. Her findings were similar to those of Ramirez and colleagues, that is, approximately 95 percent of the interactions in the regular and ESL classrooms were teacher-initiated and called for short answers that displayed rote recall of facts.

As much as we may decry this situation, we cannot fall into the trap of blaming teachers. Teachers teach in the way that they have been trained. The findings above call for changes in the way that teachers are trained, where the content of training is conveyed through methods that teachers will be expected to use, that is, interactive methods that provide students with opportunities to test language hypotheses, to express themselves, and to develop critical thinking skills as opposed to simply recalling facts. It is imperative that those who conduct both preservice and in-service programs for teachers begin to focus on the need for interactive teaching methods and on conveying training content through the use of those methods. Since teachers teach in the way they've been taught, they must be taught in the way they should teach.

Parent Involvement and Language Issues

A second area in which Dr. Ortiz calls for the assessment of schools in reference to LM students is the degree and nature of parent involvement in the schools. Although there are many cultural issues related to involving LM parents, findings from two recent studies have implications for this issue.

The first of these findings was incidental findings of Ramirez and his colleagues in the study mentioned earlier. This group found that the greatest amount of parent involvement with children's schooling was in the late-exit bilingual programs where native language instruction was used during a considerable portion of the time. There was less such parent involvement in the early-exit and structured immersion programs. Authors of the study suggest this may be due to the fact that parents in the late-exit program, where more of the instruction was provided in the native language, were better able to understand both the language of their child's instruction and the school's expectations for parent and child.

Findings from another recent study also raise cogent questions concerning the relationship among parents, families, and schools and highlight the need for additional research. Wong Fillmore (1991) and a group of volunteer researchers surveyed 3,000 families of LM children who were in all-English early childhood programs. Parents reported that not only were their children losing their first language, but this loss created consequent disruptions in parent-child interactions and relationships because parents and grandparents could no longer communicate with the children.

These two studies point to the need for further examination of language factors that mediate involvement of LM parents in their children's education. I challenge researchers to tackle the host of questions that arise in this regard. Clearly, additional research on parent involvement with respect to possible language mediation is called for.

Cultural Relevancy and Learning

A third characteristic of schools that adequately meet the needs of LM children is cultural relevancy. Three major areas where cultural relevancy affects the school experience are teacher and student interactions, curriculum and materials, and classroom management as related to teaching structures.

With regard to teacher and student interactions, teachers need to be trained in the many varieties of cross-cultural interaction and communication styles and must be cognizant of the cultural bases of their own individual styles. They also need to learn about interaction styles in the specific cultural groups represented by their students so that interactions with the teacher can be meaningful to the children and misunderstandings reduced. We all know stories of children who have been unknowingly rejected by a teacher because of a mismatch of communication styles, or children who have been encouraged and stimulated to academic productivity because a teacher knew how to use culturally appropriate means of encouragement.

Curriculum and curricular materials are another means of providing culturally relevant education. Cultural relevancy can be introduced in curricular materials through effective techniques of multicultural education. This does not mean adding on units about specific cultures so children in the classroom who represent a minority culture feel like they're being put under a microscope and their differences magnified. Instead, this means infusing multiple perspectives, or the viewpoints of several cultural groups in every possible aspect of the curriculum. For example, when studying history and current events, viewpoints of all participants in the events should be presented and examined by the students in the light of cultural beliefs. An example that has recently been in the forefront of teaching news is the controversy surrounding the discovery of North America and the perspectives that lead people to accept or reject the notion that the continent was "discovered." The notion that the continent was discovered is the perspective of only one group of people. Another example illustrates presentation of multiple perspectives at the preschool level. Although not ostensively a cultural topic, it definitely presents a different perspective. This is the children's book that recounts the story of the Three Little Pigs as told from the viewpoint of the Big Bad Wolf.

Culturally appropriate classroom management and structuring of activities can also influence learning as is demonstrated in the classical research of the Kamehameha Institute in Hawaii (Au and Mason, 1981). In that research, changing the structure of the classroom by using learning groups and activities that were consistent with the way children learned at home was associated with substantial increases in the reading scores of the students.

In assessing the appropriateness of the education provided for LM students in our schools, consideration of teacher-student interaction, curriculum and curricular materials, and classroom structuring must certainly be considered. If education is to be improved for LM students in our schools, most certainly these aspects of cultural relevancy will need to be improved.

Further Reduction of Referral and Placements

Once inappropriate referrals to special education have been reduced through improvement of the educational programs offered to all LM students, inappropriate referrals and inappropriate placements can further be reduced through specific procedures that ensure everything possible is done to solve an individual child's educational problem without referral to and placement in special education. In this regard, Dr. Ortiz has outlined and suggested detailed procedures that draw upon both her own work and relevant literature.

Although the pre-referral procedures outlined by Dr. Ortiz are comprehensive and include a number of features designed to prevent inappropriate referrals of specific individuals, it is noteworthy that these suggested procedures are all outside the realm of special education. Within the regular educational program or classroom, all possible efforts must be made to identify and resolve a LM child's problem. Only when such efforts have not succeeded should a child be referred for special education assessment. Once such a referral is made, a comprehensive assessment process should then pinpoint the source of the child's problem and determine whether special education placement is warranted.

An important characteristic of the pre-referral process to be implemented in the regular education program is team collaboration of regular education personnel, including bilingual and ESL personnel, to assist the child's teacher in identifying specific problems and in devising appropriate intervention strategies. Although employment of intervention strategies has always been a recommended procedure, the lack of collaboration and assistance provided to a teacher has resulted in a restricted range of intervention options implemented for short periods of time that usually fail to make a difference. In examining more than 1,000 special education records of LM students (Willig, Wilkinson and Polyzoi, 1985), my colleagues and I repeatedly noted this paucity of attempted alternatives and the lack of adequate documentation for the trial time periods. The procedures outlined by Dr. Ortiz, which include input and support for the teacher of personnel from regular programs in collaboration with ESL and bilingual teachers, are bound to improve the probability that problems will be resolved for individual children without progression to special education referral and placement. Conversely, the probability of identifying those students who truly need special education services and the provision of appropriate placements will be enhanced.

In sum, inappropriate referrals to special education could be reduced to a minimum through the combination of improved and appropriate regular education programs for LM students and improved pre-referral procedures to resolve specific problems. Concurrent to the attempts to accomplish this goal, an additional problem must be tackled -- the nature of instruction provided for LM students in special education classrooms.

Improving Instruction for LM Students in Special Education Classrooms

Dr. Ortiz' data indicate a reduction over time in achievement and intelligence test scores for LM children who had been placed in special education. Such data attest to the need for improvement in the nature of instruction provided to these children. Among the many problems that have been observed and documented in special education classes for LM children is the amount of individualized instruction that is actually provided, the types of teaching methods and strategies used, and inappropriate language instruction and lack of access to ESL and bilingual programs.

Although LM children are often referred to special education in hopes that they will receive more individualized instruction, several research studies have found that they may actually receive less individualized instruction in the special education classroom than if they had stayed in the regular classroom. These problems stem from the fact that frequently, pupil-teacher ratios are no different in special education classrooms than in regular classrooms and there is usually greater heterogeneity in the special education room. Resource rooms often contain students from three or four different grade levels at one time with only one teacher and sometimes an aide. Self-contained classrooms may be even more heterogeneous because of the aggregation of children who are diagnosed as mentally retarded, emotionally disturbed, learning disabled, and even physically disabled with visual or auditory handicaps. Add to this a group of children with a range of language proficiency levels in two languages and it becomes extremely difficult for the special education teacher to attend to individual needs in an effective manner.

Examples of the lack of individualized instruction for LM students in special education were evident in research I conducted several years ago with Jana Swedo (Willig and Swedo, 1987). In this study, classroom observations of LM students in special education were videotaped and analyzed for the level of task engagement under different instructional conditions. In one instance, we observed a child's individual reading session with the teacher for a ten-minute period of instruction. During eight minutes of that time allotment, the child waited in silence while the teacher attended to the many interruptions that occurred from others in the classroom. The result is that the individual reading instruction amounted to less than two minutes for that child!

To improve classroom instruction for LM students in special education, classroom management strategies must be examined and improved along with the conditions that precipitate management problems. Additionally, overcrowding special education classrooms with LM students will improve only when the assessment and improvement of regular education, as discussed earlier in this paper, occurs with consequent reductions in inappropriate special education placements.

In addition to reducing conditions that precipitate problems that limit availability of individualized instruction, improvement in special education for LM students requires an examination and adaptation of the nature of instruction offered to these students.

Sometimes changes in instruction must be radically different from the traditional task-analysis based instruction in which special education teachers have typically been trained. I observed an example of such an extreme change and the results it produced in one special education classroom of fifth and sixth grade Hispanic students. The observed student was a fifth grade Hispanic boy who had been placed in special education. In his first five years of schooling, this child had never written anything other than his name.

During the observations I made in that classroom, the teacher was experimenting with process writing as it has been described by Graves. In the first step of the writing process, children were given a story starter and asked to finish the story in any way they wanted. After completing a story, the children would read their first draft to the rest of the class, get comments, and then revise. This classroom had the highest rate and degree of task engagement of all activities observed over several months in a number of similar classrooms (Willig & Swedo, 1987). For almost and hour and a half, these fifth grade special education children were glued to their papers, writing furiously.

During the observations I conducted over a period of about six weeks, the one child mentioned, who had never written anything in the first five years of school, wrote one sentence as a story. When called upon to do so, he stood up and very haltingly read this sentence to his classmates. The students in the class had been instructed to tell each author what they liked about each story and to ask questions for clarification. When this child heard several others say there was something they liked about his sentence, he got so enthused that, by the following week, when I again visited the classroom, he had expanded the one sentence to a paragraph. By the third week, he had again read to his classmates, received more feedback, and expanded his story to a whole page of original writing! Of course, the spelling and other surface features left a lot to be desired, but this was the first time in the five years of schooling of this child that he had written anything other than his name.

The point of this is that attempts to modify instruction to produce substantial changes in outcomes will most likely require more than the minor types of modifications that teachers have been used to making at the pre-referral stage.

In summary, special education will be able to serve LM students effectively only when inappropriate referrals and placements are reduced through general improvements in regular education programs that preclude the need for many referrals, and through adequate pre-referral strategies such as those outlined by Dr. Ortiz, that reduce inappropriate referrals of specific children. Furthermore, for those LM students who truly need special education services, there is need for change in the nature of the services and instruction that is provided so that these more specifically address the language and learning needs of these children.

References

Au, H., and Mason, J. (1981). Social organizational factors in Learning to Read: The balance of rights hypothesis. Reading Research Quarterly, 17(1), 115-152.

Dale, P.S. (1976). Language Development: Structure and Function. Holt, Rinehart and Winston.

Ramirez, J. D., Pasta, D. J., Yuen, S. D., Billings, D. K., Ramey, D. R. (1991). Final Report: Longitudinal Study of Structured English Immersion Strategy, Early-Exit and Late-Exit Transitional Bilingual Education Programs for Language-Minority Children. Washington, D.C.: U.S. Department of Education.

Ventriglia, Linda. Personal communication, August, 1991.

Willig, A.C. & Swedo, J.J. (1986, April). Improving teaching strategies for exceptional Hispanic limited English proficient students: An exploratory study of task engagement and teaching strategies. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Willig, A.C., Wilkinson, C.Y., & Polyzoi, E. (1985). Interim report for the development of teaching strategies for use with LEP children who are receiving services for LD and MR. Austin, TX: Handicapped Minority Research Institute.

Wong Fillmore, L. (1991). When learning a second language means losing the first. Early Childhood Research Quarterly, 6, 323-346.

Response to Alba Ortiz's Presentation

Sherry R. Migdail
COMSIS Corporation, Mid-Atlantic MRC

I'm glad Ann told you a spelling story. I had not planned one but since this is Washington, DC, and it is a Washington spelling story, I'll toss it in.

This has to do with a new teacher whom we hired for a Washington private school. She had never lived in this city and knew little about its demography. I was in her classroom the first morning as she had asked and she was giving a spelling pretest to find out what the kids could really do when I came in to her second grade. She was very traditional and she said to them, "I will say a word, use it in a sentence, repeat the word, and then you may write the word on your paper." And she started with two or three words, the students were following her directions and things were going along well. Finally she said, "lawyer -- my father is a lawyer" and before she could repeat the word, "lawyer," 19 of the 20 youngsters looked up and said with some amazement, "Your father is a lawyer -- too?"

I don't know how many of you may have stayed up last night to see the remarkable interview with Gorbachov and Yeltsin with Peter Jennings. This is a good time to talk about the Russians. The interview reminded me of something from Anna Karenina. Tolstoy says, at some point in the book, all happy families resemble one another, but each unhappy family is unhappy in its own way. I think we can transpose this to children who are having difficulty in classrooms. Each has difficulty in his/her own way.

Many years ago when my family moved back to Washington following a number of years of living in Mexico City, our older daughter was seven years old. We registered her at a local suburban school in second grade commensurate with both her age and her previous schooling. Her English, at the time, was heavily accented although she had a fair oral knowledge of the language, her schooling had been in Spanish, for the most part.

She had been at school for but two weeks when I received a rather frantic call from the principal.

"Your daughter," she said, "cannot read."

"Cannot read -- what?" I asked.

"Cannot read what we ask her to read -- a second grade book!"

"In what language?" I asked.

The exasperated women exploded on the other end of the phone.

"Can't read in English -- that's what we teach!"

"Is that all? Try her in Spanish -- she reads quite well for a seven year old!"

"What good is that," she cried -- "we only know English."

Then she added something that was key to thinking then and still is now, "and," she said, "you are not a Spanish family -- are you?"

Our name, as you can see is not Gonzales or Rodriguez and we were not expected to know anything but English.

Well -- of course in a very short crder Lori was reading in English as well as in Spanish, but it took some doing for the principal to be convinced that she would and that she was not to be placed "back" in first grade! Testing would have determined that she was "limited English proficient" but fortunately for us the term had not yet been coined. She was not even strictly a minority language student since her family was in no way "minority." But the principal clung to the idea that she needed another "year" to become English proficient.

We all know that many children who come to school with a language other than English are for that reason overage in grade in this country.

Some years ago when I worked for the equivalent of "Head Start" in Mexico, a Guarderia Nacional for all children of ministry employees, the task for the summer was to teach a course in assessment methods and to devise or adapt instruments suited to the needs of that country. They were interested in both psychological evaluations and in a set of evaluations which could help determine possible learning problems. There were several adaptations of the Wechsler Scale for Children, which many of you know very well. Since there was no standardization of the Scale, I was asked to bring with me the Psychological Corporation Wechsler translated in the United States. I gave the test in Spanish to a youngster who had no connection with the center but he lived in the neighborhood. He was a bright, easy to talk to child and rapport was established very quickly. My purpose in giving the test was to demonstrate the futility of direct translation and the even greater dilemma in assessment when normative data is not based on a representative sample. In this case, the test was translated in the United States and distributed in Mexico.

One of the comprehension questions was very well translated; the words in Spanish exist and the translation is possible.

"Why is it better to give money to an organized charity than to a street beggar?"

The boy listened patiently, and with a kind of quizzical expression on his face asked what an "organized charity" was. I made every effort to explain but the concept of organized charity was not within the ken of this child -- nor is it a well defined concept in Mexico. Again he listened until he felt he understood. He took me gently by the arm and led me to the window. And he said, gently, "...you mean better than to give the money to my mother?" There she was with several of his younger siblings -- begging.

I brought a 16 year-old child from a village in Oaxaca to live with us in Maryland some years back. She was the daughter of our housekeeper, and I felt that her mother needed one of her children with her and had promised that when my older ones were in college and there was room, she could come.

She arrived on a Friday and by Monday I had an appointment with the teachers in her high school. The meeting was a professional courtesy and I was grateful. She sat between her mother and me and I explained to the group that she had village schooling. She is from Telistlahuaca and for any of you who know southern Mexico it is about a couple of days burro ride from the capital. When you get to the village and ask for her grandmother's house, you are told that it is "under the Pepsi Cola sign." You can't miss the sign, but if you do you are out of the village.

I had known the child most of her life and she always appeared to be a bright and capable person. We had her grades and they were well within the average range. I explained the 1-10 grading system and described her school and something about how she was taught.

It was the counselor who broke the ice.

She stood over the child and spoke in a loud and clear voice beginning with, "IF I SPEAK SLOW--LY YOU WILL UNDERSTAND ME." The child, seated between her mother and myself, elbowed us both as she hissed between her teeth, " Y esa loca...quien es?" (And this nut - who is she?)

And one other story from my perspective. I was in La Paz, Bolivia, at the San Andres University doing a presentation to a group of students and teachers. The topic -- second language acquisition and the implication for classrooms. Many of you know that there are two important languages in Bolivia, Aymara and Quechua. The bilingual issues have to do with getting Aymara and Quechua children comfortably into Spanish speaking classrooms. At one point during the discussion period, one of the young professors stood up to ask a question. It became a minispeech. "Was the estimable doctora aware," he started, "that here in Bolivia there has been a considerable body of research and experience related to the Aymara and Quechua people and that there is unrefutable evidence that because of brain structure which is different for these persons than for Spanish speaking Bolivians, it is now proven, that learning Spanish is not possible for them." And with this startling statement he took to the board and drew a crude representation of the brain and while he drew he commented about how, because of certain formations in the brain structure, Aymara and Quechua people would not succeed...would not learn...and it was useless to try.

"Was the good doctor aware of this ongoing research work and could I comment?"

Before I had a chance to even get the astonished look from my face, a gentleman from the back of the room spoke in almost hushed tones. His anger was overriding despite the restrained tone of voice. He spoke in beautiful Spanish:

He began.

"Siendo Aymara..." "As an Aymara, I need to make certain things clear. When I was a child, they came to my village from I think the ministry of education. They came with books and with "tests" and they had all us answer questions. They spoke in Spanish and we in Aymara and there was no way we children could answer their questions. They came away calling us dullards -- it was then that I knew what I had to do...and I can assure you that I have not swayed from my mission. I am at this university to be sure that Aymara and Quechua children no longer have to be "dullards." What happened to me will not happen to my children."

I can still feel my reaction of that moment. I never answered the professor's question -- it was answered far better than I could have done.

Why do I start my comments this way? They are not amusing stories -- but they are real and they happen in one version or another everyday and in many places.

In each case, one very different from the other, the children were not behaving as the school wanted them to -- my child was not programmed in English and, therefore, from the viewpoint of the school, she needed another year in first grade until she met the "standard." Perhaps then she would be "grade appropriate" for content. The Mexican child obviously gave the wrong answer about how organized charities and his mother were related. You and I know the problem was that I gave the wrong test. The child from a Oaxacan village who knew no English was treated in a most demeaning manner! And she knew it! And my Aymara friend went through what we know continues to happen not only a continent away but also in this country when children are put into inappropriate placement because they can't pass the test. We know an inappropriate test will give you inappropriate results.

Dr. Alba Ortiz redefines pre-referral intervention in her paper. It is her feeling that the traditional framework may be too narrow and she redirects pre-referral as having two major components:

First, "a prevention component aimed at establishing educational environments conducive to the academic success of language minority students so that problems will not occur in the first place," and second, "a problem-solving component in which the teacher first adapts instruction and/or the classroom environment to improve student performance and then requests assistance from others if problem solving efforts are not successful."

In her comprehensive paper, she also elaborates on the phase from referral to assessment to placement. I especially agree with the need for collaborative/school community relationships, especially with the parents of the children. Call them once when a child has a good day and you will have made a friend for life -- for you and for the child. Obviously, cultural and linguistic incorporation in the curriculum means a whole lot more than hanging a pinata in the middle of the room as you convince yourself you've done your bit for the Hispanic children. The use of interactive approaches to language minorities is essential.

It is interesting that many years after our second stint in Latin America our third child told us a story I can't forget. Remember we are not Hispanics...we lived abroad and brought home to this country children who were Spanish speaking -- at that point -- Spanish dominant! It seems that our young Karen, then about seven, was in the hall in an excellent suburban school and overheard two teachers, both hers, talking about a trip the class was to make. They were talking about what benefits were to be derived from the excursion when one said, "...all but Karen. It's too bad she doesn't understand much -- you know her family speak Spanish." Where were the high expectations teachers know we need for successful schooling for a Spanish-dominant child?

I might add that it took years before Karen really liked school!

I want to briefly make mention of Dr. Ortiz's emphasis on students who may be inappropriately placed for years in special education on the basis of a poorly planned assessment. Dr. Ortiz indicates in her paper:

"...after three years of special education placement, Hispanic students who were classified as learning disabled had actually lost ground. Their verbal and performance IQ's were lower than at initial entry."

The paper also reviews the characteristics of "empowerment" pedagogy stressed by Jim Cummins. I would like to extend those characteristics and apply them to assessment procedures.

- genuine dialogue between student and teacher

Ask the right questions -- assume the individual you are speaking with doesn't need a loud voice, or sign language, but recognize that the student's language can be assessed as you get him to talk with you. In writing, try a dialogue journal, emphasize process writing, build a "portfolio." Keep a record of his work, chronologically and thereby see a pattern of growth with the student alongside. Show and Tell! Tell him he can do it and show him where he has made progress. If some of this sounds familiar better deja vu than to be marking time in place.

- encouragement of student-student talk in a collaborative learning context

- focus on developing higher level cognitive skills rather than factual recall. To do this, you must recognize that a language minority student has the ability and capacity for higher level cognitive skills.

I was quite concerned when Dr. Ginsburg felt that he did not have a clue as to why there was a disproportionate number of language minority kids in special education. He felt that one important research question should be, "...what are the characteristics of kids in special education?" We know those characteristics....I want to know the characteristics of the teachers who put the kids there in the first place! They need help!

If we were to use the model of interactive pedagogy as a basis for interactive assessment then we assess children not parts of children and language, not in its small bits -- but as a whole.

One can even make some pretty accurate "guesses" in interactive assessment about kids who are limited English but who have a good "sense" of language. One gets a feeling about intonation and rhythm. One can appreciate the functionality of language; the effectiveness of a student's ability to communicate. Does this kid get his message across? To what extent? Does he have an inner ear and hear himself and does he begin to make corrections. This metacognition is a very important feature of second language learning in a very practical sense. When you learn a second language you hear it in your head and if it doesn't "sound right" you try to fix it before you actually say a word or phrase.

And let's not stop there. You also get a good feeling for affect and for risk-taking behaviors, for motivation, for anxiety.

Pre-referral, bearing both Ortiz components in mind, prevention and problem solving, eliminates much of the disorientation if and when teachers are given the guidance they need. I suggest that in the Ortiz context of "referral teams" the best "team" is a group of teachers who see the student in a variety of contexts -- physical education, music, classroom, ESOL and so on. Build in the notion that diagnosis is for improvement of instruction not for finding remediative procedures.

I am presupposing a system where there is, to use Dr. Ortiz's phrase, "a collaborative learning community on the school's campus."

But I need to also talk briefly about the student who, despite all efforts, will and does experience difficulty. Dr. Ortiz's paper discusses clinical teaching, which, with the best of skill and intentions, does not always work. Often, it is not just help with reading. If we agree that a percentage of youngsters have neurologically-based learning problems, then this population will also have its share.

In this geographic area, Washington DC, and environs, there are between 60 and 80 languages spoken by the students in public schools. The largest number is Spanish speaking. Others include Chinese, Vietnamese, Kymer, Loatian, Urdu, Hindi, Gjuarati, Portugese, Swedish, Croatian, Polish, Russian, all middle European languages, Greek; even Yap, Chomorro, Hausa, Igbo and Sango.

One of our major local school systems began a team approach for bilingual assessment in 1980. I was its founding member. Over the years the process has been refined and the team has been expanded; bilingual interpreters for a number of languages, a bilingual consultant psychologist, a bilingual speech and language therapist, and counselors are available.

We were, in 1980, concerned about the language minority child who was "suspected of being handicapped" and for whom an assessment might be indicated. Looking back at those early years I am convinced that we were on the right track. Our team was responsible for working with the teacher initially, for gathering data and developmental, social, and educational histories; for classroom visits; for meeting with parents and talking with them about their expectations for their children, finding the right question to ask and, finally for assessing the student and making recommendations.

We did workshops for the schools, including for the speech and language people, Head Start teachers, and other specialists, and we talked with school administrators about individual children and about our work in general. We looked for trends -- and we found them.

There were too few of us -- and too many of them! The needs and the demands on our time were very great. I am certain they still are. About five years later we did some internal research. Numbers of students we had seen, ages, gender, their time in the country and in the schools, their grade level at the time of the initial referral, parental information, and so forth.

Our hypothesis was that of the group of close to a thousand children who had been assessed by the Team, the smallest number would be designated as needing some form of "special education." We were right. Of that special education needs group, we found a number of youngsters who were learning and/or language disordered. The largest number was in need of extra attention.

We found that despite good oral skills many of the children were being referred by fourth to seventh grade teachers. Logically it was because the students were an enigma despite good oral skills, "He knows English as well as I" kind of syndrome. Reading in English was difficult -- many read but did not comprehend easily or comfortably. Writing skills were even less well developed. These students were not disabled -- they needed additional help. Any number of these kids were not being recognized for what they could do. I remember children who were undoubtedly gifted or talented but unrecognized. I remember children who were bored to tears, overage in grade...and I remember confused parents.

Learning disabilities is an American concept...it really is. Other countries have rushed onto the bandwagon but they have not yet confided the l.d. phenomenon to parents -- certainly not parents in rural schools in Salvador or Guatemala. They come here -- hard working people who want to improve the lot of their children -- and are told that the child they brought from Matahualpa or Esquintla who functioned pretty well at home is "disabled."

We must train our teachers to appreciate the essence of "cultural difference." It is of vital importance to know something about where people come from and what "disability" may denote in other cultures.

Let me conclude with two "cultural" stories. The first is lovely and touching and certainly a tribute to hardworking and dedicated ESL and Bilingual teachers. The second is a firm illustration of what we need to know but may always be afraid to ask!

A group of us were in a meeting with an Ethiopian parent whose six children were in a local school. An explanation was given for the child's problem and a suggestion was made that the boy be "temporarily" put in a special class. He was academically below grade -- at least a couple of years. He was an Amharic speaker and English was just beginning to make sense. He could barely read. The father had been employed at the American air base in Ethiopia and spoke English well enough not to need an interpreter. He was adamant about keeping the child, about 10 years-old, in a mainstream class. His final word was a strong and powerful argument.

"Give him a chance," he said. "I am grateful for your interest and I know you mean to help my son. But I need you to know he was born in a cave above Addis Ababa during our troubles at home and I don't care when he reads. I am grateful for his life and I know he needs time to grow."

A second story concerns a youngster, the son of a Nigerian diplomat. The child's father had the permitted number of wives: four -- and a great number of siblings. He was referred for special education but needed a psychological assessment to make the final determination. Since the boy spoke Hausa and the psychologist did not, he was asked to draw a picture of his family. This is a fairly usual procedure in nonverbal testing from which a psychologist will make a number of assumptions.

Mohammed was given a large piece of paper and a crayon and he began to draw. First a large stick figure, then four small figures. Then he counted. One...on his fingers, with his yes turned upwards, he subvocalized, one, two...and drew some five small stick beings. Again the same procedure, first the count and then some five more... a third time, count and draw. The psychologist, exasperated after the first dozen small figures turned to the boy and insisted: "I said your family -- not your tribe."

Mohammed, however, very serious at his task, very task-oriented, said, I am, I am -- I'm almost finished -- I only have one more mother to do!"

To repeat Tolstoy, all happy families resemble one another but each unhappy family is unhappy in its own way. Maybe that's why we teachers of children and teachers of teachers must really recognize cultural pluralism for what it contributes to our lives as well as to the lives of "our" children.

The Assessment of Alternative Certification Practices: Panel Presentations

Annalisa Allegro
New Jersey Department of Education

In an attempt to get through all the material, I hope that I don't lapse into my East Coast double speed talk. I want to thank Dr. Garcia this morning for his wonderful introduction of New Jersey; although I feel it may be unfounded, it was good to hear New Jersey spoken of so highly.

New Jersey's alternative certification route has been acclaimed by several organizations. One well known organization quotes -- that "the New Jersey alternative certification route is one of the most effective and promising strategies for improving teacher supply and quality and therefore, public education." Education Week recently came out with an editorial on alternative certification entitled, "Alternative Certification Is An Oxymoron," and I believe that fits the New Jersey model.

This is not going to be a typical bureaucratic presentation. What I'd like to do is to tell you about the New Jersey model and some of the pitfalls as it pertains to bilingual and ESL teachers. Alternative certification began in New Jersey in 1985. It was to be implemented for bilingual and ESL teachers in 1991 but has been delayed a year. So what I am speaking of are predictions for the future. And also, as a prelude to this, I think I need to explain -- or put alternative certification in a context, the development of it. I'd like to provide you with some background on New Jersey.

There are 567 school districts in the small state New Jersey; 410 districts have LEP students. We have mandates for bilingual education. If there are 20 or more LEP students of a single language group in a district, bilingual education programs must be instituted. Bilingual education programs always include ESL instruction. There are more than 80 bilingual programs in New Jersey. Spanish programs are the most predominant, but we also have Arabic, Japanese, Korean, Haitian Creole, Polish, Vietnamese Mandarin, and Gujarati. So we are in need of many bilingual teachers and we are in need of teachers from many different language backgrounds. There are also more than 170 ESL-only programs. Therefore, there is a great demand for ESL teachers in the 80 bilingual programs as well as in the 170 ESL-only programs.

Alternative certification has been proclaimed as a way of increasing the pool of teachers. New Jersey provides certification for both bilingual and ESL teachers. The regulations for certification are developed by the Department of Education and passed by the State Board of Education. Prior to the current changes, to become a bilingual teacher, you needed to have an elementary or a content area certificate plus a bilingual endorsement which was an 18-credit hour course of study. An ESL certified stands by itself, covers grades K through 12, and was a 24-credit hours course of study until changes were instituted.

As part of the background it is important to know that the Department of Education in New Jersey is separate from the Department of Higher Education and, specifically, the division of teacher certification developed all certification rules without much input from the office of bilingual education.

Alternative certification is simply an alternate way of becoming a teacher without completing a preservice college program. Alternative certification has three areas: formal instruction, school-based supervision, and evaluation. Sounds pretty good, formal instruction so you get the theory; school-based supervision, with mentorship and coaching sounds great; and lastly, evaluation. Let's take a look at it for bilingual and ESL teachers, bearing in mind that we wanted to increase the pool and improve the quality of teachers.

In order to become eligible to be an alternative route candidate, you have to have a bachelor's degree in some area -- history, science, whatever. You must then pass the NTE Communications Skills Test. If you want to be a bilingual teacher, you have to pass an NTE subject area test. At the high school level, you would have to have a subject area test -- science, history, or social studies. At the elementary level, you would need to pass the NTE General Knowledge, Test which poses great difficulty for language minority candidates. Now we have two standardized tests that a candidate must pass to get into the alternate route. No candidate can enter a classroom until both tests have been successfully passed.

The provisional program is generally a one-year program but, for bilingual and ESL teachers, it turns out to be a two-year program because general education courses and the specific courses in bilingual education or ESL are required. If you already have a teaching certificate in New Jersey, you would be exempt from the 12-15 credits of instructional theories in education and curriculum, learning development, and classroom management.

To become an ESL teacher, you would take those courses in the first year and then, in your second year, you would take 180 hours of classes in linguistics, second language acquisition, methodology, etcetera, which works out to be 12 credits. Prior to this, 24 credits were required. To become a bilingual teacher, you would take 200 hours of general education courses and then, in your second year, 90 hours in bilingual education. That works out to be six credit hours. Prior to the changes, 18 credits in bilingual education theories, methodologies, etcetera were required. I also would like to add that this formal instruction of general education and bilingual ESL is not provided by universities, it is to be provided by state centers. We have had no development of such state centers to date for bilingual and ESL education. The district also has the option of providing this training.

While we've begun to collaborate with the school district, we haven't provided enough support to implement such professional training programs.

Let's move to the aspect of supervision. Here is where a team of people help the teacher candidate through his or her first year of teaching. One of the people in this group is designated a mentor, an experienced teacher, who would supervise, assist, and train the teacher candidate. Now that sounds great! Let's take a look at it.

The support team consists of the principal, the school, a mentor teacher experienced in the area, if possible, and two other staff in related fields, for instance, a curriculum person or a content area department chair; During the first 20 days as a teacher candidate you are not fully responsible for the classroom. A mentor teacher is in the classroom and oversees the candidate's teaching. Thereafter, for the next two and a half months, you are to be observed once a week by one of the members of the support team. For the next three months, there are at least monthly observations. Now let's look at who the support team is going to be. In large bilingual programs such as Newark, they have a hard time hiring Bengoli and Gujarati teachers. Who is qualified to mentor the Gujarati candidate? Within the large Spanish or Portuguese bilingual programs they are unable to borrow any of the certified bilingual teachers to mentor new candidates. Let's take a look at a small suburban Spanish bilingual program with only two bilingual teachers. It's virtually impossible to mentor a candidate because both teachers' schedules are filled.

Now let's look at ESL teacher candidates in a district with a new program. Who is going to supervise the ESL teacher candidate? There are no certified ESL teachers there. The state says that any experienced teacher could supervise. An elementary teacher could mentor a teacher candidate for bilingual education. A history teacher could mentor a teacher for the secondary bilingual teacher candidate. That's fine, but do they have experience or training in working with LEP Students? The same thing is true with an ESL teacher. While any certified teacher would be acceptable to the state, what kind of quality supervision, training, and mentorship are you providing for these teacher candidates?

There must be an evaluation of teacher candidates three times during the year. Two evaluations are formative, with the first one after 10 weeks of being in full charge of the classroom, the second formative evaluation after 20 weeks. The mentor teacher would provide these evaluations and the principal would have input on the summative evaluation. How is the mentor teacher going to evaluate the teacher candidate on his or her ability to communicate, to provide appropriate responses or appropriate lessons if the mentor teacher does not understand the language and has not worked with LEP students?

In order to increase the pool of teachers, incentives should be provided for districts or teacher candidates. The teacher candidate pays a fee of $450 to the mentor teacher, $550 must be paid to the experienced teacher who serves as a member of the support team, and a fee of $600 must be paid to the Commissioner of Education for instruction provided at the regional center. The teacher candidate is responsible for these fees. While the teacher candidate is paid a salary, many teacher candidates start at the minimum which is $18,500. I don't know if there is much incentive for people to enter the field of bilingual/ESL education in New Jersey.

There's one other aspect I would like to add to this presentation. In New Jersey there is no requirement to receive any other education beyond a bachelor's degree while teaching. You are not required to take additional university credits or in-service credits. When I taught in the District of Columbia, six hours of in-service were required every five years. Many teachers in New Jersey do take courses, of course, but there is no mandate for continuing education. The quality of the teaching force is questionable.

There is something to learn from the New Jersey experience. The basic format of how it was developed is a framework but actual implementation and state support must be strengthened. The state of New Jersey has issued many mandates, therefore, it has great control over our education. In this case, the department of education has blocked its commitment to quality education for second language learners, by quality education, you are not going to be increasing the pool of teachers by issuing regulations which stops the opportunities for additional bilingual and ESL teachers.

Migdalia Romero
Hunter College, New York

The three of us decided -- we didn't collaborate, we knew what our topic was, alternative certification, alternative forms of certification, and each of us developed something I think complements each other's presentation. Mine really deals more with a framework for certification. I've divided my presentation into five parts: first I'm going to state the problem; second, give you a framework, state of the art and where things need to be going; third, give you five interesting practices that come from the field that exemplify alternative forms of certification; fourth, address some unresolved issues in the alternative framework for certifications; and fifth, end up with some recommendations.

A colleague of mine -- he was also a Title VII Fellow -- actually he's a faculty member in the Department of Curriculum in Teaching, where I am a chair at Hunter College, did his dissertation on Puerto Rican males, 180 Puerto Rican males. I'm going to share one statistic with you because I think it exemplifies part of the problem I'm going to be addressing. The statistic is that he found the average GPA of the Puerto Rican male population that he was working with was 2.42. The average GPA of the white population that was part of the group was 2.97.

Now, the Hunter College teacher education program has just raised its GPA for entry into the teacher education program from 2.5 to 2.7. That forecloses -- closes out -- some of the very people that we need to be attracting into the field of teacher education and, in particular, bilingual education. Other kinds of things that are happening in the field, for example, using Hunter as a reference point, is that we have so many students we're trying to get into the TESOL program that we have closed enrollment for matriculation for one semester. Again, in a field where we need people -- this is a public education institution that's supposed to be preparing people for working in public schools -- we're closing doors.

The new certification and recertification processes that are being used throughout the country and, in particular, the exams that are being used to certify people as teachers are weeding out the language minorities and the very people that are needed for our bilingual program. I see the problem as threefold, and I'm going to address each of these. First, the minority representation problem; second, there is a testing problem; and third, there is a problem with how certification is implemented, how it is perceived and the philosophy behind it. In the first problem of representation, we find an under-representation of the students that are being served by the public schools. So, if we have large Hispanic populations or Asian populations in the schools, we don't have teachers in sufficient numbers to serve these students. I'm not suggesting that only Hispanic teachers can teach Hispanic students or only Asian teachers can teach Asian students. But Asian teachers, Hispanic teachers, Haitian teachers, bring in knowledge, sensitivity, and skills that you can't easily develop in a teacher training program.

As an example, I cite the statistics I got from AACTE, 16.2 percent of all school youths are African-Americans, and only 10.3 percent of teachers are African-American. Nine percent of the students are Hispanic, but fewer than 2 percent of teachers -- this is the National Education Association, 1987, statistics. The projection for the future is more dismal in that our most talented minorities are not coming into teaching. There's not enough of an incentive economically for them to do so, they are seeking other fields. So, the recruitment problem is very real. There are more opportunities for minorities in other fields and there is poor recruitment of them. That's the representation problem.

The second problem is the testing problem. There is an increased reliance on testing as a means of certification. In fact, as of April 1987, 48 states had adopted some form of testing, but only seven had included satisfactory performance observation. I say that because the direction I'm going to be moving is looking at performance as part of that certification process. As part of the testing problem, there is a higher fail rate of minorities in certification tests. From 1986 to 1987, 81 percent of white candidates passed California's state certification test and, 34 percent of blacks; 59 percent of Mexican-Americans failed and 51 percent of other Latinos. So we're seeing, again, the group of teachers that we're trying to attract are doing least well on paper and pencil tests, and we need to talk about that a little bit more.

To give you another example of how tests are weeding out people and the problems that are created by tests, I cite statistics on the New York State teacher candidates taking the National Teacher Exam (NTE) in March of 1990. In communication skills, 37.8 percent of Asian students passed; Puerto Ricans, 48.5. In general knowledge, I'm just going to point to the lowest statistic, the Puerto Ricans, who in general knowledge saw only 37.4 percent pass. Overall, 70 percent passed, so half that number were passing among Puerto Ricans. In professional knowledge, Puerto Ricans again scored the lowest pass rate, 52.3 percent, Asians were 53.6 percent, and whites were 89.3 percent. Quite a difference.

There are clearly some issues and problems that are suggested by these statistics, and I think one of the recommendations I'll be moving toward is we need more discrete analysis of the tests that are given to find out what the issues are, what the problems are, and what we need to be addressing in teacher training if we are to attract some of the people that we need. The tests, in fact, are keeping out some of the minorities whose sensitivities, cultural and linguistic knowledge, and skills, and their performance in classrooms, would most benefit English language learners. Another example, and it's just a personal example of a teacher when I first started teaching. She was a wonderful, wonderful kindergarten teacher. In fact, I chose her class to put my daughter in. My daughter at that time was four or five years old. This teacher took the NTE at least four times and had tremendous difficulty passing it. An exceptional teacher, a teacher who was perfectly fluent in English, born and raised in New York, and fluent in Spanish but more fluent in English, but she had trouble passing the NTE.

So, we're going to see it over and over in some of the comments I'm making that the tests have poor predictability. The test often does not test the skills that are needed in classrooms with children. It tests accumulative knowledge of students, and we're not sure of the relationship between that knowledge and the effectiveness of that individual in a classroom. The question is, how do we improve the certification process so that we are not relying exclusively on tests to weed out individuals who could be good teacher candidates?

The framework that I'm going to speak of -- I've divided it into -- first of all, looking at a distinction between training and certification. Certification is a process, the way it is perceived now and the way it is acted on now is a process that is evaluative in nature, it is a stamp of approval by an agency, a state education agency or a local education agency, a stamp of approval that this person is qualified to teach. It is also private. You do it, usually a paper and pencil test, you do it alone, and it is a process that screens out individuals.

Training, on the other hand, is a supportive process; it is a reflective process. It is established to engage people in thinking through what they are doing, in planning, in evaluating, what they're doing in a classroom and how effective they are. It is interactive in nature. Teachers in a training program interact with students, they interact with colleagues in a professional way and it leads toward improvement.

The reason I make that distinction between certification and training is because in more enlightened states or districts, the field and the direction of the field seem to be going toward the melding of certification with training, where training becomes part of the certification process. Alternative certification was defined by one of our previous speakers as a way to circumvent preservice education programs and to allow teacher candidates to go directly into schools and to help them get certified.

Again, the implication is that a training process that is labeled as an alternative certification process is a way of getting teachers to become better at what they do, taking people without any education background and getting them to a point where they are certified or endorsed. Traditionally, there have been two points at which teachers are certified, preservice certification, which qualifies them to go into a classroom and do practice teaching, usually at the end of a bachelor's program with some credits in education, and then at the end there is a stamp of approval that says they are fully qualified, and we call that [illegible] in-service or post-service certification process. One is for temporary certification and the other is for permanent certification.

The field is moving toward mastery certification or recertification. There's a group called The National Board for Professional Teacher Certification. In fact, there is a meeting next week, and I'm going to be part of that group that is looking at certifying teachers at a national level, so we're not talking state or local certification. We are talking about certification for teachers for purposes potentially of merit pay, of being able to take that certification to any other state, and it gives them a lot more flexibility in seeking jobs. So the field is moving toward a mastery certification process.

What are some of the routes to certification that currently exist? The first one I keep alluding to is the test -- using a test as the basis for certification. It is supposed to be an objective test. The second route to certification is certifying -- and states often do this, they will certify a program, an institution of higher education submits a plan, this is how we certify our teachers, these are the credits they have to take in these different areas, this is the amount of field supervision they have, and the state certifies the program. And if teachers go through -- and all of you are familiar with that -- they come out certified.

Currently 49 states accredit teacher education through a process known as program approval. That is, using state standards, institutions design preparation programs that are subsequently approved by the state. However, the certification process is only as good as the programs, and the teachers are only as good as the programs they go through.

The third route to certification is neither the test nor the program, but it is again the movement of the field toward performance-based certification. Let me just list a few of the formats that seem to be taking: videotaping and self-evaluation, observation of teachers, portfolio maintenance of student work, peer review, mentorship, coaching, interviews with teachers to find out what it is they are doing professionally. There seems to be a movement in the field toward greater on-site supervision required as part of the alternative certification process. Basically, what we are certifying is either experience, accumulated knowledge, or demonstrated skills. I have a problem with that because I think we also need to be looking at creativity as part of that certification process, the degree to which individuals are able to reflect on what they do and improve on what they do.

The way tests stand now they cannot test creativity. So one element of a really good teacher is being able to be reflective about what he or she does and improve on it. Tests don't test that. So, this is why the field is moving toward certification as an ongoing process, where you can look at creativity as part of that process.

Let me share with you some practices in the field, mentorship programs; we have one at Hunter College. It's a four-year support program. New teachers are placed in schools and assigned to mentor teachers, but there's also a university faculty member who goes into the classroom, observes teachers, and does demonstrations when necessary. It's a very supportive program. New Jersey will be talking about another mentorship and alternative certification program that it has.

Another example is an International High School in New York that gets its teachers much more involved in certification process through a supportive certification process in which new teachers are paired off with more experienced teachers and they put together portfolios. The portfolios include work that they have done as teachers, lessons they have done that they think are exemplary. They can select their own lessons -- student work that exemplifies the kinds of experiences they have given their kids and how their students have grown. In the portfolio one can do documentation of professional development, conferences attended, etcetera, as well as observation reports, logs, and self-analysis of how information on students has been used to improve or change teaching. The documentation is on creativity and reflection.

By 1992, even Educational Testing Service (ETS) is moving in this direction. ETS will replace the national teachers' exam which is most frequently used for teacher certification with a comprehensive teacher assessment profile including computer simulations, interactive video, portfolio development, and classroom observation. ETS believes that comprehensive assessment administered at different points in a perspective teacher's education will give students a better chance to demonstrate the knowledge and skills that relate significantly to classroom performance.

What are my recommendations? Number one, we need more discrete analysis of the needs of minorities in the certification process and of their strengths so as to build on those strengths, to tap those strengths and to meet the real needs. Secondly, we need a comprehensive teacher certification process of which testing is only one part and a process that uses multiple assessment methods. Third, we need a melding of certification and training so that the process of developing teachers is used as part of that certification process. Fourth, we need more careful examination of and attention to supervised field experience as part of the certification process. Fifth and finally, we need to bring the certification process in line with our thinking about teaching as an intellectual and creative art.

Elena Izquierdo
District of Columbia Public Schools

Confronted with serious budgetary constraints, changing demographics in the city, policies that lowered the mandatory school age (age 5), new federal and state requirements for disabled learners and, in addition, competition with surrounding school districts both in recruitment and retention, the District of Columbia was challenged to look for very creative and innovative ways of refining the roles of its teachers in addressing the needs of the students in the district. The District of Columbia Public Schools initiated a Retooling Initiative in order to meet the needs of its student population.

The Retooling Initiative was aimed at certified teachers with classroom experience and demonstrated competency with the goal of reequipping them to perform new roles.

Teacher Retooling Survey

The District of Columbia developed a Teacher Retooling Survey to begin the process of utilizing its existing resources in meeting the needs of its student population. The teacher survey was developed and disseminated to all teachers in the DC schools in May 1991. It asked teachers whether they were interested in retooling in the areas of critical need for its student population: Bilingual Education, ESL, Early Childhood, Special Education, Bilingual Special Education, Science, Mathematics, Elementary Education, Foreign Languages, Computer Science and Technology. Teachers were also asked to give their current certification area, teaching assignment, years of experience, language proficiencies, and degrees. It is important to note that this was a voluntary survey, and teachers who responded did so only because they were interested. The office of Research and Evaluation for DC Schools then compiled all the data and presented each division with its respective data.

The results of the survey were phenomenal! There were over seven hundred (700) teachers who expressed an interest in the area of Bilingual/ESL.

In the District of Columbia, there are more than 9,000 language minority students representing over 100 different language groups.

The number of language minority students with limited-English language proficiency increases daily in the district. In an effort to meet the needs of limited-English proficient students, and maintain within the existing framework of limitations, the Language Minority

Affairs Branch, Local Education Agency for Bilingual Education, used the results of the teacher survey and ventured into a Bilingual/ESL Teacher Retooling Institute.

Selection Process

With a response of more than seven hundred (700), we began to organize a very rigorous process for teacher participant selection given that I only had the funds to retool fifty (50) teachers. One of the questions on the survey asked teachers whether they would be willing to begin classes in the summer of June 1990, with six (6) credit hours, if funding was made available to them. Not all of the seven hundred (700) teachers interested were able to commit. This provided us with the first cut in the selection process.

These teachers were then contacted for an orientation meeting that provided them with an overview of the critical needs of the district, Bilingual/ESL Education, the Retooling Institute, and an outline of the areas of competencies in Bilingual/ESL needed for certification (Historical, Philosophical, Educational, and Sociological Bases of the Education of Language Minority Students; Understanding the process of First and Second Language Acquisition; Methodologies, Learning Styles; Multicultural Education; Alternative Assessments; Principles of Effective Instruction for the Education of the Language Minority Child). Teachers had to commit to the entire summer program and, in addition, commit to course work in the fall, spring, and a practicum during the following summer for a total of twenty-four (24) graduate credit hours leading to certification in this area. In addition, it was explained to them that, at the end of the course work and certification, they would be placed in schools where there was a need for services in the area of Bilingual/ESL Education. They also had to commit to work in this field in the DC schools for a minimum of two (2) years. Their commitment to the Retooling Initiative was strongly emphasized, and they were discouraged to apply if they could not commit. This provided the second cut of possible retooling applicants.

Those teachers who expressed their commitment were then given an application to complete. The application consisted of information such as teaching experience, degrees held, past and current assignments, language proficiencies, and certification areas. In addition, they were required to write a 250-word essay on why they were interested in retooling in Bilingual/ESL Education. They had to include with their applications a recommendation letter from their current school principals and copies of their performance evaluations for the last three (3) years. The turn around date was two (2) weeks. Once received, the applications were reviewed for complete information and documentation. Those teachers who met the application requirements were notified of a specific date and time for a Panel Interview. Two panels were organized in order to interview all applicants. The panel interviews were conducted simultaneously every twenty (20) minutes for two (2) days.

The panel participants consisted of principals, Bilingual/ESL teachers, and administrators knowledgeable in the field of Bilingual/ESL Education, for a total of four participants for each panel. Each participant had a score sheet for each applicant. Questions regarding methodologies used with language minority students, cultural sensitivity, interest in retooling and commitment were asked. Although the panelists recognized the fact that the applicants had little if any training in these areas, their sensitivity and responses were important. Again applicants were asked about their commitment to the Retooling Initiative, which meant that summer vacation took on a different meaning for them, and fall meant going to classes after school and/or on Saturdays. In addition to this, they were reminded that, upon completion of their coursework, they would be placed within a school that needed Bilingual/ESL teachers for a period of at least two (2) years. After each interview, panelists were instructed to complete the scoring sheet along with added comments. Based on every step of the application process, required documentation, and the interview, the teachers for the Retooling Institute were selected. I cannot over emphasize the importance of the selection process. Interested and committed applicants went through the entire process. Upon review of all documents, the teachers were selected for the Bilingual/ESL Retooling Institute.

Bilingual/ESL Teacher Retooling Institute

The Bilingual/ESL Retooling Institute was a collaborative venture between the District of Columbia Public Schools and the George Washington University. Joel Gomez, Director of the National Clearinghouse for Bilingual Education, and Dr. Alicia Martinez, professor for the George Washington University, Teacher Preparation Department, were instrumental in developing the institute and coursework leading not only to certification and its direct application into the classroom but also to the overall success of the Retooling Institute. The goal of the institute was to retool mainstream teachers in Bilingual/ESL Education and prepare them for teaching language minority students, particularly limited-English proficient students. In the summer of 1991, DC mainstream teachers began their coursework in this area. The coursework consists of twenty-four (24) graduate credit hours inclusive of a six (6) hour practicum. The practicum includes site visits to schools with successful Bilingual/ESL programs, seminars, and teaching experience in a Bilingual/ESL classroom setting. Dr. Alicia Martinez has provided these participants with guidance, direction, and most important is her role as their mentor. This has been a critical component of the entire institute which has led to the success of the Bilingual/ESL Retooling Initiative. These retooling teacher participants have evolved into one of the most professional, wonderful, and knowledgeable groups in the field of Bilingual/ ESL Education I have seen in a while. They are working with their current principals and sharing with them and the school staff issues related to the education of language minority students.

By the completion of their course of study, the District will have teachers that are already in the system, that bring with them years of teaching experience in content areas, knowledgeable in the District's curriculum, and now knowledgeable in the education of language minority students. One of the most rewarding outcomes will be that these newly certified Bilingual/ESL teachers will have years of experience in teaching in the content areas -- and this is what the field of Bilingual/ESL Education is demanding of its teachers in order to appropriately meet the needs of language minority students. These Bilingual/ESL Teacher Retooling professionals are now some of the strongest advocates for the education of language minority students in the District of Columbia Public Schools. **The Bilingual/ESL Teacher Retooling Institute -- DC's alternative to Alternative Certification.**

Discussion of Panelists Allegro, Romero, and Izquierdo's Presentations

Barbara Clements
Council of Chief State School Officers

I'm really happy to be here today. The whole issue of alternative certification for bilingual teachers is something I've been thinking about for six years. Right after New Jersey implemented its alternative certification program, Texas implemented a program and I worked with it at the Texas Education Agency. So I have some personal experiences I'm going to include in my discussion of some of the key things that were discussed today.

I have titled my comments "the Cons and Pros of Alternative Certification for Training Bilingual and ESL Teachers." As I said, I've been thinking over this for a long time, and I definitely think there are some good things and some not-so-good things going on. I think you've heard some of the issues discussed. So, let me summarize some of the cons and pros.

Teacher training and certification requirements are frequently mentioned as being the most formidable barrier to attracting new teachers. There's a lot of discussion about the Mickey Mouse courses that you have to take in teacher education programs, and I went through one of those programs. So, I know what they're talking about. Over the last few years there has been a lot of discussion about alternative certification and it has received the support of a lot of states. Even the President came out expressing support for alternative certification as a means of attracting highly trained and knowledgeable people into the classrooms.

I know that, in the most recent literature, I have read that some type of alternative certification programs have been adopted by approximately half of the states as a means of attracting non-education majors who are either mid-career or retired or perhaps young. The idea is to attract them into the teaching profession primarily to ease teacher shortages but also as a means of attracting better educated people. Now, it's interesting to me that there has been a high level of interest concerning this. I was fascinated that more than 700 teachers in D.C. expressed an interest in possibly being retooled through an alternative certification program.

In Texas we had an enormous amount of response to the alternative education activities that were going on there. I have a 1990-91 report that indicates there were 1,242 interns in 191 districts, so there are an awful lot of people out there who are interested in getting certified if an opportunity is made available to them.

Let me refresh your memory or reiterate some of the general requirements for most alternative certification programs. First of all, the candidate must hold a bachelor's degree. Generally this means a bachelor's degree other than in education; it is possible to get an education degree and not get certified. Secondly, the candidate must pass some sort of standardized test. Some states require the standardized test to be a basic skills test, or it might be a communication skills test. In other states it might be a content knowledge test. That's the requirement in Texas. A third requirement is that there must be some sort of compressed training that occurs before the candidate actually enters the classroom, usually that training covers instructional design, measuring student performance, and other relevant topics. Again, this is prior to becoming a teacher of record.

It appears to me that most of these training activities usually cover about four weeks. It's hard to imagine having your entire teacher education program compressed into four weeks, or at least a substantial portion of it.

A fourth requirement is that during the one year on-the-job program, and I understand that some states are already making it a two-year program, around 200 classroom hours of pedagogy are taken. They are taken over a period of time, and they are generally not done through standard university courses as you've heard today. But there are a fair number of hours that are covered.

Fifth, they have some type of support. Usually a mentor is assigned to the candidate to provide individualized support, and often there is a support team to advise and evaluate the applicant or the candidate. Now, you've heard a number of problems associated with programs of this type. New Jersey has been in this business longer than anybody else, so they really know where the potential problems are, even though they haven't trained bilingual teachers yet.

First of all, the presenter from New Jersey mentioned inadequate training, and again, I stress that we're talking about compressing an awful lot of content, knowledge, and pedagogical knowledge into a short amount of time. There are many things that we know are effective, such as cooperative learning and effective classroom management, things that are really important to getting off to a good start, having good student cooperation throughout the school year, and also learning.

Also the content that persons in alternative certification programs may have had in their college careers may not be age appropriate, and it may not reflect current theories or understanding in the area. So, there is a need to refresh their memories about the content, particularly if they're mid-career or retired people. A second type of problem has to do with the mentor teacher. Frequently the mentor teacher may not be in the same content area because there may not be a teacher on hand in a school who's teaching exactly the same content, or maybe it's not the same language. It's very hard for the mentor teacher to evaluate, to know how to evaluate and do a good job of evaluating teacher candidates; particularly if they're not within content areas it's hard to give constructive feedback. It's also more difficult to share materials -- the collaborative piece that is supposed to be important to alternative certification.

Other problems with mentor teachers include what if they are not in the same school, or what if one mentor teacher has a whole lot of people assigned to him/her. Some programs have up to 20 people assigned to one mentor teacher. What about the fact that the mentor teachers frequently are not given any training or expectations as to what they're to do. They have to kind of "wing it."

Finally, what is the support for or incentive to participate in an alternative certification program? Some states or some school districts are offering some additional money. That's important because it reinforces the teacher for participating. I think there is an expectation that once a career ladder is established, that's a logical role for a higher level teacher to play.

A third set of problems have to do with evaluation. I noticed in one program described today that the first formative evaluation doesn't take place until the end of the first 10 weeks. That's a long time for a brand spanking new teacher to work in a classroom without any fairly structured feedback. I think new teachers of all sorts need continual support and feedback, and they also need to be taught to self-evaluate, and I'm not sure that's being included as a part of these programs.

A fourth item I want to mention has to do with passing tests because it was mentioned and because I was associated with the testing program in Texas. We did find a smaller number -- fewer of our minority candidates were passing the test than the Anglo test takers. On the other hand, of the people who were participating in the alternative certification program, a higher percentage of them are passing the test than the general teacher education school population. So it is interesting that it tends to depend on who gets brought into an alternative certification program.

Now I notice that many people who complete alternative certification programs are hired to fill vacancies in rural and urban areas. As you know, a lot of these are the most difficult areas to try to teach in. I have observed some of these classrooms and they are unbelievably difficult. Some of the bilingual classes have maybe five or six different languages within the same classroom. But it's interesting. I read that in New Jersey they've found that people who go through the alternative route tend to leave the profession less often than those who go through the traditional preservice education route. It's also interesting to me that a large number of minorities are being recruited through these alternative certification programs.

I was involved in doing some interviews in Texas of the people who were participating in our programs. Many of them said to me that really they always wanted to be teachers. The minority teachers, in particular, said that they were encouraged to go into other fields because now the opportunities exist. Once they got into those fields and were working as chemists or physicists or mathematicians they found that they really kept thinking about what it would be like to go and teach. This included a lot of people whose parents had been teachers or school administrators. So I think there is a potential for bringing into the field a number of people who would have preferred to be there from the beginning but just got steered otherwise.

Now, there's a difference in alternative certification programs for elementary and secondary teachers. Most of the programs that the states currently have are geared toward secondary teachers. The assumption is, if you've got a degree in a content area that you can come into a secondary classroom and teach. In my personal experience, having observed in many, many classrooms and having done some teaching myself, I feel that probably at the secondary level this could work. You could probably get enough information, or nearly enough information, in those compressed four weeks of training to get off to a fairly good start, and with the 200 hours over the school year you might be able to do a fairly passable job during your first year of teaching and be able to benefit from the training and do a much better job thereafter.

However, I have some questions about the utility of these kinds of programs for elementary teachers and particularly for teachers of LEP and disabled children. It appears that in some of these programs the things that we expect of people going through preservice programs are not necessarily required for the people who are coming through alternative certification programs. I'm not so sure that's a good thing to do. So I think there is some potential there for improving the quality of the instructional requirements for teachers coming through the alternative route.

Let me close by asking the question, do I think that alternative certification programs can adequately produce bilingual teachers and address the shortages that are being felt by Texas, California, and D.C., and other places. I think perhaps they can. I think they could help. I think it depends however on what they do and what they must do. So to address what alternative certification programs can do, it's important to consider what happens to traditionally trained teachers. Teachers who come through traditional programs enter into employment with a set of skills and knowledge that they received in college. The assumption is made that they are sufficiently well-trained to start teaching on the very first day of school. They bear equal responsibility with all the other teachers. With some exceptions, most districts give these teachers virtually no assistance during their first year of teaching. Generally, in-service training is focused on administrative requirements.

But research on beginning teachers indicates that they need a lot more assistance at the beginning of their careers. They need help with curriculum planning, classroom management, and evaluation of students, and most beginning teachers think they are not getting enough of that in their preservice education. Alternative certification programs on the other hand are generally run by districts, and in order to meet the requirements of an acceptable program, the district must carefully plan the program to ensure the candidates are adequately trained and they get the support they need to succeed.

So as a result they can tap into resources not traditionally tapped, such as teachers, and they can bring in outside consultants to help them. These programs focus on providing candidates with the kind of one-on-one training, the collaboration, the spirit of cooperation, the kinds of things that beginning teachers indicate they need during their first year of teaching. So in that way alternative certification programs may be providing beginning teachers with something that our regular beginning teachers, the teachers going through traditional programs, are not getting.

Then, to address what must be done by these programs in order to succeed, I'll quote AACTE requirements for what they think are necessary components of a program for alternative certification. They say that new candidates must receive information about child and adolescent development, measurement of student performance, information on recognizing student handicaps, legal rights of students, and finally, and I think most importantly, the impact of cultural diversity on learning styles.

So can we adequately train bilingual teachers through these programs? I don't think the results are in yet. Now if through these programs we have better access to well-educated, native speakers, we might be improving the quality of our bilingual teaching force since these people have the background and the knowledge of the language, and they can promote native language proficiency as well as English language proficiency. So they may be giving our bilingual children a better shot at succeeding, and our monolingual non-English speaking children as well. If the commitment of the teachers trained through alternative certification programs is stronger, and we have some evidence because they do tend to be more mature than those who go through traditional programs, we may also be retaining more of our bilingual teachers for a longer period of time. And we may substantially reduce the shortage in this particular area.

We must not, however, stop monitoring the quality of the program and their capacity to meet the needs of teachers who are going to be working with this very special group of children.

Teachers for Language Minority Students: Evaluating Professional Standards

Eugene Garcia
University of California, Santa Cruz

Introduction

The policy debate regarding the education of language minority students in the United States has centered on the instructional use of the native and/or the English language as a medium and/or target of instruction. For educational professionals and educational researchers, the more specific issue of concern has become the identification, implementation and evaluation of effective instruction of a growing population of ethnolinguistic minority students who do not speak English and, therefore, are considered candidates for special educational programming that takes into consideration this language and cultural difference. Research on this issue has involved representatives of psychology, linguistics, sociology, politics, and education in cross-disciplinary dialogue. For a thorough discussion of these issues see August and Garcia (1988), Baker and deKanter (1983), Cummins (1979), Garcia (1983), Garcia (1991), Hakuta and Garcia (1989), Hakuta and Gould (1987), Ramirez, Yuen and Ramey (1991), Rossell and Ross (1986), Toike (1981), Willig (1985). The central theme of the discussions is the specific instructional role of the native language. At one extreme of this discussion, it is recommended that the native language play a significant part in the non-English-speaking student's elementary school years, from 4-6 years, with a set of standard of native-language mastery prior to immersion into the English curriculum (Cummins, 1979). At the other extreme, immersion into an English curriculum is recommended early, as early as preschool, with minimal use of the native language and concern for English Language leveling by instructional staff to facilitate understanding by the limited-English-speaking student (Rossell and Ross, 1985).

Each of these disparate approaches argues that its implementation brings psychological, linguistic, social, political, and educational benefits. The *native-language* approach suggests that competencies in the native language, particularly as they relate to academic learning, provide important psychological and linguistic foundations for second-language learning and academic learning in general -- "you really only learn to read once." Native-language instruction builds on social and cultural experiences and serves to politically empower students from communities that have been historically limited in their meaningful participation in majority educational institutions. The *immersion* approach suggest that, the sooner a child receives instruction in English, the more likely he or she will be to acquire English proficiency -- "more time on task, better proficiency." English proficiency in turn mitigates against educational failure, social separation and segregation, and, ultimately, economic disparity. Such a debate has clearly affected the type of educational professional which should serve these students.

As this debate developed during the 1970s and 1980s, it became clear that the students who came to school speaking a language other than English received considerable attention in research, policy development, and practice. The Department of Education and the Department of Health and Human Services, as well as private foundations, supported specific demographic studies and instructional research related to this population of students, preschool through college. The United States Congress authorized legislation targeted directly at these students on five separate occasions (1968, 1974, 1978, 1984, and 1988), and numerous states enacted legislation and developed explicit program guidelines regarding both instructional alternatives and the requirements of educational professional who would be allowed to serve these students. Moreover, federal district courts and the U.S. Supreme Court concluded adjudication proceedings that directly influenced the educational treatment of language minority students.

The intent of the present discussion is not to focus on the ongoing debate, but instead to utilize the data generated by that debate to assess our present understanding of who the students are that language minority teachers are serving, what types of instruction these students are presently receiving, and, most significantly what types of teachers are presently serving these students. A major presupposition of this discussion is that "who" does the teaching is of major significance regardless of the language minority education model which is being implemented. The discussion will also attempt to extend the data base by cautiously but directly addressing future directions with regard to the development of "effective" language minority teachers. Of particular concern will be credentialing policies and their political and empirical underpinnings. The overall purpose of this discussion is to suggest ways in which to enhance the educational plight of language minority students by focussing on the educational professionals who directly serve these students on a daily basis. A much more localized district level teacher evaluation/ credentialing alternative is prepared for evaluating language minority teachers.

Defining Language Minority Students

The search for a comprehensive definition of the "language minority student" reveals a variety of attempts. At one end of the continuum are general definitions such as "students who come from homes in which a language other than English is spoken." At the other end are highly operational definitions such as, "students who scored in the first quartile on a standardized test of English language proficiency." Regardless of the definition adopted, it is apparent that students vary widely in linguistic abilities. The language minority population in the United States continues to be linguistically heterogeneous. Not inconsequential is the related cultural attributes of these populations of students, which are not only *linguistically* distinct but also *culturally* distinct. Describing the typical language minority student, therefore, is highly problematic. In simple terms, the language minority student is one who (a) is characterized by substantive participation in a non-English-speaking social environment, (b) has acquired the normal communicative abilities of that social environment, and (c) is exposed to a substantive English-speaking environment, more than likely for the first time, during the formal schooling process.

Estimates of the number of language minority students have been compiled by the federal government on several occasions (Development Associates, 1984; O'Malley, 1981). These estimates differ because of the definition adopted for identifying these students, the particular measure utilized to obtain the estimate, and the statistical treatment utilized to generalize beyond the actual sample obtained. For example, O'Malley defines the language minority student population by utilizing a specific cutoff score on an English language proficiency test administered to a stratified sample of students. Development Associates estimates the population by utilizing reports from a stratified sample of local school districts. Therefore, estimates of language minority students have ranged between 1,300,000 (Development Associates, 1984) and 3,600,000 (O'Malley, 1981).

In 1976, the total number of language minority children aged 5-14 approximated 2.52 million, with a drop to 2.39 million in 1980 and a projected gradual increase to 3.40 million by the year 2000 (Waggoner, 1984). In 1983, this population was more conservatively estimated to be 1.29 million (Development Associates, 1984). In 1983, this population was more conservatively estimated to be 1.29 million (Development Associates, 1984). This divergence in estimates reflects the procedures used to obtain language minority counts and estimates. These children reside throughout the United States, but distinct geographical clustering can be identified. About 62 percent of language minority children are found in Arizona, Colorado, California, New Mexico,and Texas (Development Associates, 1984; O'Malley, 1981; Waggoner, 1984). Of the estimated number of language minority children in 1978, 72 percent were of Spanish Language background, 22 percent were of Asian background, and 1 percent were of American Indian background. However, such distributions will change, due to differential growth rates, and by the year 2000 the proportion of Spanish language background children is projected to be about 77 percent of the total (O'Malley, 1981). Estimates by Development Associates (1984) for students in grades K-6 indicate that 76 percent are of Spanish language background; 8 percent, Southeast Asian (e.g., Vietnamese, Cambodian, Hmong); 5 percent, other European; 5 percent, East Asian (e.g., Chinese, Korean); and 5 percent, other (e.g., Arabic, Navaho). For national school district sample in the 19 most highly impacted states utilized by Development Associates, 17 percent of the total K-6 student population was estimated to be language minority in these states.

Regardless of differing estimates, a significant number of students from language backgrounds other than English attend U.S. schools. As this population increases steadily in the future, the challenge these students present to U.S. educational institutions will increase concomitantly.

Educational Programs Serving These Students

For a school district staff with language minority students, there are many possible program options: e.g., Transitional Bilingual Education, Maintenance Bilingual Education, English-as-a-Second Language, Immersion, Sheltered English, and Submersion (General Accounting Office, 1987). Ultimately, school staffs reject program labels and focus instead on the following questions: (a) What are the native language (L1) and second language (L2) characteristics of the students, families, and communities to be served? (b) What model of instruction is desired? This involves the question of utilizing L1 and L2 as mediums for instruction as well as handling the actual instruction of L1 and L2. (c) What is the nature of the school and resources required to implement the desired instruction?

Programs for language minority students can be differentiated by the ways they utilize the native language and English during instruction. A report by Development Associates (1984) was based on a survey of 333 school districts in the 19 states serving over 80 percent of the language minority students in the United States. For grades K-5, they report the following salient features regarding the use of language(s) during instruction: (a) 93 percent of the schools reported that the use of English predominated in their programs, and conversely, 7 percent indicated that the use of the native language predominated; (b) 60 percent of the sampled schools reported that instruction was in the native language and English; (c) 30 percent of the sampled schools reported minimal or no use of the native language during instruction.

Two-thirds of these schools have chosen to utilize some form of bilingual curriculum to serve this population of students. However, about one-third of them minimized or altogether ignored native language use in their instruction of language minority students. Programs that serve Spanish-speaking background students have been characterized primarily as *Bilingual Transitional* education. These programs transition students from early grade, Spanish-emphasis instruction to later grade, *English-Emphasis* instruction and eventually to *English-Only* instruction.

Recent research in transition type programs suggests that language minority students can be served effectively. Effective schools organize and develop educational structures and processes that take into consideration both the broader aspects of effective schools reported for English-speaking students (Purkey & Smith, 1983). Of particular importance has been the positive effect of intensive instruction the native language that focuses on literacy development (Wong-Fillmore & Valdez, 1986). Hakuta and Gould (1987) and Hudelson (1987) maintain that skills and concepts learned in the native language provide a basis for acquisition of new knowledge in the second language.

For the one-third of the students receiving little or no instruction in the native language, two alternative types of instructional approaches, English as a Second Language and Immersion, predominate. Each of these program types depends on the primary utilization of English during instruction but does not ignore the fact that the student served is limited in English proficiency. These programs are used in classrooms in which there is not a substantial number of students from one non-English-speaking group. These programs have been particularly influenced by recent theoretical developments regarding second-language acquisition (Chamot & O'Malley, 1986; Krashen, 1982), and indicate that effective second-language learning is best accomplished under conditions that simulate natural communicative interactions.

It is important to note that the bulk of language minority students served in today's public schools are in elementary schools. The most comprehensive data is still that of Developmental Associates (1984). They report that the schools in their national sample identified three to four times as many Grade 1 students as Grade 5 students. Moreover, 20 percent of students in grades 1 to 3 were transitioned into an English curriculum in any one year. More recent is Olson's (1989) California data which indicates that some 73 percent of language minority students are in grades K-6. Those schools sampled by Developmental Associates (1984) and a similar national sample studied by Halcon (1981) provide some empirical data with regard to the instructional staff that serves these elementary students:

1. The schools serving language minority students in grades 1-5 had 4.0 teachers, 3.5 paraprofessionals and 1.1 resource or instructional support staff (Chapter 1 aide, Migrant aide, etc.).
2. Teachers in these classrooms had a median 5.8 years of experience teaching language minority students. However, 50 percent of these teachers had less than 3 years of teaching experience with language minority students.
3. Less than 50 percent of teachers responsible for instruction of language minority students spoke a language other than English.
4. Less than 30 percent of these teachers had obtained language minority education related credentials.

This service and staffing data indicate that school district staff have been creative in developing a wide range of programs for language minority students. They have answered the previously listed questions differentially for (a) different language groups (Spanish, Vietnamese, Chinese, etc.), (b) different grade levels within a school, (c) different language subgroups of students within a classroom and even different levels of language proficiency. The result has been a broad and, at times, perplexing variety of program models. It is also clear that these programs are staffed extensively with paraprofessionals and with teachers who have limited teaching experience with the population of students they serve, with half not able to speak the student's native language, and with more than two-thirds not holding a specific professional credential related to language minority education.

Effective Teachers for Language Minority Students

Although it is difficult to identify specific attributes of teachers that have served language students effectively, recent efforts have attempted to do so. Unlike earlier reports which have identified and described effective programs, recent efforts have sought out effective programs and/or schools, then attempted to describe the specific instructional and attitudinal character of the teacher (Carter & Chatfield, 1986; Garcia, 1988; Garcia, 1991; Pease-Alvarez, Garcia and Espinosa, 1991; Tikenuff, 1983; Villegas, 1991). This new emphasis on the language minority education teacher is related to the broader interest in identifying "exemplary" teacher characteristics for teachers in general (Reynolds and Elias, 1991). Dwyer (1991) identifies four domains which "good teachers excel in: (1) content knowledge; (2) teaching for student learning; (3) creating a classroom community for student learning; and (4) teacher professionalism. Villegas (1991) has extended these four domains when the student population served by the teacher is culturally and linguistically diverse. She suggests that "good" teachers in these classroom contexts are required to incorporate culturally responsive pedagogy. To go beyond these generalizations, the following section describes specific research which has attempted to document empirically the attributes of effective language minority teachers. These studies are few, but they begin to provide a set of practice standards which may be useful in training and evaluating language minority teachers.

A concern for the effectiveness of teachers is not new. From the earliest days of education program evaluation, the quality of the instructional staff has been considered a significant feature (Heath, 1982). Unfortunately, for programs serving language minority students, the evaluation of "effectiveness" has been consumed by an empirical concern regarding the significance of the use/non-use of the students' native language and the academic development of the English language (August and Garcia, 1988). Very little attention is given to the attributes of the professional and para-professional staff which implements the myriad of models and program types omnipresent in the service of language minority students. Typically, attention to the characteristics of such a staff is restricted only to the years of service and extent of formal educational training received (Olsen, 1988). Yet, most educational researchers will grant that the effect of any instructional intervention is directly related to the quality of that intervention's implementation by the instructor(s).

Attention to "exemplary" programs and "exemplary" teachers comes from the great dissatisfaction the field of language minority education has come to realize with regard to the limited conclusions and unproductive debates regarding the relative effectiveness of bilingual education (Hakuta, 1985; Hakuta and Garcia, 1989). This field has continually been subjected to national evaluations. The most recent is the Ramirez, Yuen, Ramey and Pasta (1991) study, which attempts to assess the academic effects of various bilingual, ESL, and other approaches. Such studies are continually criticized for their methodological flaws, and, have little effect on the field—on what teachers do in classrooms (August and Garcia, 1988). Beginning with Tikunoff (1983), more in-depth studies of "good" language minority schools and classrooms addressed the specific organizational and instructional characteristics in programs which were "working" for language minority students. Such an emphasis suggests that there is much to learn from programs that are serving language minority students well. Instead of searching for the "best" model by doing large scale comparative studies, all which will likely be methodologically flawed, this new line of inquiry suggested that we search out effective programs and carefully document the attributes which make them effective. From such data, other programs seeking to better serve language minority students could at least compare themselves to these "exemplary and effective" organizational features, instructional practices and teacher attributes (Carter and Chatfield, 1986; Garcia, 1988; Pease-Alvarez, Garcia and Espinosa, 1991 and Garcia, 1991).

It is in this more "micro" spirit, that the present discussion attempts to specifically advance our understanding of what makes "good" language minority teachers. Such a discussion requires the reliable identification of the "exemplary" teacher, no small task, along with the interview and observation of these individuals. In addition, interviews of school administrators and parents should assist in a more comprehensive perspective of these significant individuals. It is not the purpose of this discussion to suggest that all "good" language minority teachers need to be like the ones described in the present literature. Instead, it is the intent of the discussion to carefully describe the attributes of these effective teachers in such a way that others may make use of this information to better serve language minority students.

Tikunoff (1983), in his report of the Significant Bilingual Instructional Features (SBIF) study, reports commonalties in the "exemplary" teacher's response to organization and instruction of classrooms. The 58 teachers observed in this study covered six sites and included a variety of non-English languages. All classes were considered effective on two criteria: First, teachers were nominated by members of four constituencies -- teachers, other school personnel, students, and parents -- as being effective. Second, teaching behaviors produced rates of academic learning time (a measure of student engagement in academic tasks) as high as or higher than reported in other effective teaching research.

An initial set of instructional features identified for the effective teachers pertains to the delivery and organization of instruction:

1. Successful teachers of limited-English-proficient (LEP) students specify task outcomes and what students must do to accomplish tasks. In addition, teachers communicate high expectations for LEP students in terms of learning and a sense of efficacy in terms of their own ability to teach.

2. Successful teachers of LEP students, not unlike effective teachers in general, exhibit use of active teaching behaviors found to be related to increased student performance on academic tests of achievement in reading and mathematics including: (a) commu-

nicating clearly when giving directions specifying tasks and presenting new information; (b) obtaining and maintaining students' engagement in instructional tasks by pacing instruction appropriately, promoting involvement, and communicating their expectations for students' success in completing instructional tasks; (c) monitoring students' progress; and (d) providing immediate feedback whenever required regarding students' success.

3. Successful teachers of LEP students mediated instruction for LEP students by using the students' native language and English for instruction, alternating between the two languages whenever necessary to ensure clarity of instruction. Although this type of language switching occurred, teachers did not translate directly from one language to another.

The SBIF study also reports that the teacher made use of information from the LEP students' home culture so as to promote engagement in instructional tasks and contribute to a feeling of trust between children and their teachers. The SBIF researchers found three ways in which home and community culture was incorporated into classroom life: (a) Cultural referents in both verbal and nonverbal forms were used to communicate instructional and institutional demands; (b) instruction was organized to build upon rules of discourse from the L1 culture; and (c) values and norms of the L1 culture were respected equally with those of the school.

In more recent research which focused on Mexican-American elementary school children, Garcia (1988) has reported several related instructional strategies utilized by effective teachers. These teachers were nominated by language minority colleagues and served students who were scoring at or above the national average on Spanish and/or English standardized measures of academic achievement. Garcia's (1988) research characterized instruction in the effective classrooms as follows:

1. Students were instructed primarily in small groups and academic-related discourse was encouraged between students throughout the day. Teachers rarely utilized large group instruction or more individualized (mimeographed worksheets) instructional activities. The most common activity across classes involved small groups of students working on assigned academic tasks with intermittent assistance by the teacher;

2. The teacher tended to provide an instructional initiation often reported in the literature (Mehan, 1979; Morine-Dershimer, 1985). Teachers elicited student responses but did so at relatively non-higher-order cognitive and linguistic levels; and,

3. Once a lesson elicitation occurred, teachers encouraged students to take control of the discourse by inviting fellow student interaction, usually at higher-order cognitive and linguistic levels.

Teachers in the Garcia (1988) study fulfilled general expectations reported by Mehan (1979) for regular expectations and by Ramirez (1986) and Ramirez, Yuen, Ramey and Pasta (1991) for language minority teachers. Teachers did not invite instructional interaction in other than the most communicatively simple mode (factual and truncated "answer giving"). This type of elicitation style may be particularly problematic for Hispanic Language minority students in that these students may not be challenged by this style of instructional discourse to utilize either their native or second language to express complex language functions which reflect higher-order cognitive processes. However, teachers were clearly allowing student-to-student interaction in the child-reply component of the instructional discourse segment. Teachers encouraged and engineered general student participation once the instructional peer interaction was set in motion. This finding is particularly significant. Garcia (1983) suggests that such student-to-student interaction discourse strategies are important to enhanced linguistic development. Wong-Fillmore and Valadez (1986) report that peer interaction was particularly significant for enhancing second language oral acquisition in Hispanic children. Moreover, Kagan (1986) has suggested that schooling practices which focus on collaborative child-child instructional strategies are in line with developed social motives in Mexican American families. The interactional style documented in this study seems to be in concert with that which is most beneficial, both linguistically and culturally, to Mexican American students.

A recent study (Garcia, 1991) focused on three teachers, a first grade, third grade, and fifth grade teacher, in a highly regarded Spanish/English, bilingual school. These teachers were consistently identified at the school site level and at the district level as "effective" teachers. Approximately 50 percent to 70 percent of their students were Spanish dominant, the remainder were English dominant. The findings of this study with regard to teacher attributes were divided into four distinct but interlocking domains: (a) Knowledge, (b) Skills, (c) Dispositions, and, (d) Affect.

Knowledge

These teachers were all bilingual and biliterate in English and Spanish. They had the prerequisite state teacher credentials and had graduated from specific bilingual, teacher-training programs. They had an average of 7.1 years experience as bilingual teachers. Therefore, these were not novice teachers with little general teaching or language minority teaching experience. In addition, they reported that they routinely participated in staff development efforts, either taking courses or attending workshops on techniques that they wanted to implement in their classrooms. Some of the workshops, sponsored by the school or district, were mandatory. These teachers also participated in courses that they sought out and financed on their own, some related to Spanish language development and others related to pedagogy and curriculum.

These teachers were quite knowledgeable and articulate with regard to the instructional philosophies which guided them. They communicated these quite coherently in their interviews. They never hesitated in addressing "why" they were using specific instructional techniques and usually couched these explanations in terms of a theoretical position regarding their role with regard to teaching and "how" students learn. Principals and parents also commented on these teachers' ability to communicate effectively the rationales for their instructional techniques. One principal commented, "She's always able to defend her work with her students. When she first came here, I didn't agree with all that she was doing, and sometimes I still do not agree. But she always helps me understand why she is doing what she is doing. I respect her for that. She is not a 'recipe teacher'." A parent commented with regard to her children's journal writing: I didn't understand why she was letting _____ make all these spelling mistakes. It annoyed me. During the teacher-parent conference, she showed me the progress _____ was making. His spelling was getting better without taking a spelling test every week. I was surprised. She knows what she's doing." A parent concerned about his daughter, not competent in English in the third grade, indicated, "Me explicó que aprendiendo en español le va a ayudar a mi hija hablar mejor el inglés. Dice bién, porque mi hijo que vino conmigo de Mexico, hablando y escribiendo en español, aprendio el inglés muy facil." Moreover, these teachers seemed to be quite competent in the content areas. The upper elementary teacher who was instructing students in fractions had a solid and confident understanding of fractions. She did not seem to be "one step ahead of the students."

Skills

Despite their differing perspectives, the teachers demonstrated specific instructional skills. They used English and Spanish in highly communicative ways, speaking to students with varying degrees of Spanish and English proficiency in a communicative style requiring significant language switching. Direct translation from one language to another was a rarity, but, utilization of language switching in contexts which required it was common.

Of course, variations existed among these exemplary teachers. However, each had developed a particular set of instructional skills which they indicated led to their own effectiveness:

1. Teachers had adopted an experiential stance toward instruction. Along with many of their colleagues, these exemplary teachers had abandoned a strictly skills-oriented approach to instruction. To varying degrees, they organized instruction in their classes so that children first focused on that which was meaningful to them. Early grade teachers used an approach to reading instruction that treated specific skills in the context of extended pieces of text (e.g., an entire book, passage, or paragraph). They initiated shared reading experiences by reading to and with children from an enlarged book, pointing to each word as they read. Because most of these books relied on a recurring pattern (e.g., a repeating syntactical construction, rhyming words, repetitions), children who could not read words in isolation were able to predict words and entire constructions when participating in choral reading activities. With time, teachers encouraged students to focus on individual words, sound-letter correspondences, and syntactic constructions. The teacher also encouraged children to rely on other cueing systems as they predicted and confirmed what they had read as a group or individually.

These teachers also utilized a thematic curriculum. Science and social studies themes were often integrated across a variety of subject areas. Once a theme was determined, usually in consultation with students, the teachers planned instruction around a series of activities that focus on that theme. For example, a unit on dinosaurs included reading books about dinosaurs, categorizing and graphing different kinds of dinosaurs, a trip to a museum featuring dinosaur exhibits, writing stories or poems about a favorite dinosaur, and speculating on the events that led to the dinosaurs' disappearance. In the third grade classroom, a student suggested that the theme address "the stuff in the field that makes my little brother sick": pesticides. The teacher developed a four week theme which engaged students in understanding the particular circumstances in which many of them reside with regard to pesticide use.

Despite the use of instructional strategies that depart from traditional skills-based approaches to curriculum and instruction, these teachers did sometimes structure learning around individual skills or discrete components. For example, the teachers devoted a week or two to preparing students for standardized tests. During this time they taught skills that would be tested and administered practice tests: "I don't like testing. But we have to do it. I teach my kids how to mark the bubbles and I make sure that they take their time. We practice test-taking, but we don't take it seriously."

403

2. Teachers provided opportunities for active learning. These teachers organized a good portion of class time around a series of learning activities that children pursued either independently or with others. During science and math, children worked in small groups doing a variety of hands-on activities designed to support their understanding of a particular concept (e.g., classification, estimation, place value) or subject area (e.g., oceanography, dinosaurs).

Teachers' commitments to active learning were revealed in their commitments to a studio or workshop format for literacy instruction. Instead of teaching students about reading and writing, teachers organized their program so that students actively read and wrote. Real reading and writing took place in the context of a literature-based reading program and during regularly scheduled times when students wrote in their journals on topics of their own choosing and teachers responded to their entries. There was also time for students to engage in writers' workshops. During this time students generated their own topics, wrote, revised, edited, and published their finished writings for a larger audience. As with adult published authors, they shared their writing with others and often received input that helped them revise and improve upon what they had written. For example, one teacher commented, "These kids produce their own reading material and they take it home to share it with their parents. It's real good stuff. I help a little, but its the kids that help each other the most."

3. Teachers encouraged collaborative/cooperative interactions among students. These teachers organized instruction so that students spent time working together on a wide range of instructional activities. The two primary grade teachers structured their day so that students worked on group and individual activities (e.g., graphing, journal writing, science projects) in small heterogeneously, organized groups. Students who worked in small groups on their own art project, journal, or experiment did not necessarily interact with other members of their group. Teachers explained that students, particularly those who did not share the same dominant language, often ignored one another during these kinds of group activities. They felt that cross-cultural interactions was much more likely to take place when students were obliged to work together to complete a single task.

Dispositions

The following descriptions of teacher attributes were considered "dispositions" because no other category seems relevant. They are individual characteristics which these teachers possessed. They are likely to be relevant to their success more as professionals than as teachers. For instance, these teachers were highly dedicated. They reported working very hard, getting to school first and being the last to leave, working weekends, and sometimes feeling completely overworked. They reported spending close to $2,000 of their own resources in modifying their room and obtaining the materials their students needed. They indicated that they saw themselves as "creative," "resourceful," "committed," "energetic," "persistent," and "collaborative." They sought out assistance from their colleagues and were ready to provide as much assistance as they received.

Although these teachers felt that they were effective, they were not complacent. They continued to change their instructional practices and in some cases their instructional philosophies over the years. These teachers reported experiencing great change in their approach to learning and instruction, having shifted "paradigms." These teachers, who once advocated skills-based and authoritarian modes of instruction such as "DISTAR," are now considering and experimenting with child-centered approaches. Teachers felt that they enjoyed a certain degree of autonomy in their school. They felt free to implement the changes that they wanted. In recent years, when they have wanted to implement something new in their classroom, they have gone to their principal with a carefully thought-out rationale and have eventually enlisted her/his support. These teachers have been involved in change that has had an impact on other classrooms as well as their own. Along with other teachers, they have obtained support to eliminate teaming and ability grouping across subject areas in the first grade. In addition, they were actively involved in the district-wide teacher-initiated movement to eliminate kindergarten testing. These teachers were involved in individual and group efforts to improve the quality of education at the school and district level. In short, these teachers were highly committed to improving themselves and the services to students in general.

Above all, they were highly confident, even a bit "cocky" regarding their instructional abilities: "I have changed my own view on how students learn -- we need to understand learning does not occur in bits and pieces. Why do teachers still insist on teaching that way?" "I know what I am doing is good for kids. Some of my colleagues say I work too hard -- I say they do not work hard enough. Not that they are lazy, they just don't seem to understand how important it is to do this job right"; "I know my kids are doing well, all of them. I would rather keep them with me all day then send them to someone who is supposed to help them in their 'special' needs but doesn't help them at all."

Affect

These teachers had strong feelings that classroom practices that reflect the cultural and linguistic background of minority students are important ways of enhancing student self-esteem. These teachers felt that part of their job was to provide the kind of cultural and linguistic validation that is missing in the local community known for deprecating the Latino culture and Spanish language. According to these teachers, learning Spanish and learning about Latino culture benefits Anglo students as well as Latino students. In their eyes, people who learn a second language tend to be more sensitive to other cultures. Like other teachers, these teachers felt that being bilingual and bicultural enriched their students' lives.

Latino culture is reflected in the content of the curriculum in various ways. The two primary grade teachers, who organized their curriculum around a variety of student-generated themes, addressed cultural experiences of Latino students within the themes. For example, in a unit on monsters, they highlighted Mexican legends and folktales that deal with the supernatural (e.g., "La Llorona"). In addition, these teachers emphasized the importance of reading and making available literature that reflects the culture of their Latino students. They also encouraged students to share favorite stories, poems, and sayings that they learned at home.

These teachers had high expectations for all their students: "No 'pobrecito' syndrome here -- I want all my students to learn and I know they can learn even though they may come from very poor families and may live under 'tough' conditions. I can have them do their homework here and I can even get them a tutor -- an older student -- if they need it. I understand that their parents may not be able to help them at home. That's no excuse for them not learning." In many respects, these teachers portrayed themselves as quite demanding, taking no excuses from students for not accomplishing assigned work and willing to be "tough" on those students who were "messing around."

Most significant was the teachers' affinity toward their students: "These students are like my very own children"; "I love these children like my own. I know that parents expect me to look after their kids and to let them know if they are in trouble"; "When I walk into that classroom I know we are a family and we're going to be together a whole year....I try to emphasize first that we are a family here....I tell my students, 'You're like brothers and sisters' and some students even call me Mom or Tia. It's just like being at home here." Each teacher spoke of the importance of strong and caring relationships among class members and particularly between the teacher and the students. They felt that this provided students with a safe environment that was conducive to learning.

Parents also reported a similar feeling. They directly referred to the teachers in the interviews as extended family members, someone to be trusted, respected, and honored for their service to their children. These teachers were often invited to "bautismos," "bodas," and "fiestas de cumpleaños," and also to soccer games and family barbecues. And they attended such occasions, reporting that such participation was inherently rewarding and instructive with regard to their own personal and professional lives. Parents commented during interviews: "La señorita ________, le tengo mucha confianza, quiero que mi niño la respete como a mi"; "Nunca se larga mi nina de ella, se porta como mi hermana, siempre le puedo hablar y me gusta mucho ayudarle"; "I know my son is well cared for in her class, I never worry -- she even calls me when he does something good."

This discussion has focused on attributes of teachers who are considered "effective" for language-minority students. These teachers are highly experienced, not novices in teaching or in the instruction of language minority students. They are highly skilled in communication with students, parents, and their administrative supervisors. They think about and communicate their own instructional philosophies. They work hard to understand the community, families, and students which they serve and incorporate into the curriculum attributes of the local culture. They have adopted instructional methods which are student centered, collaborative and process oriented -- no "worksheet" curriculum here. They are highly dedicated, work hard, collaborate with colleagues and continue to be involved in personal and professional growth activities. Most significantly, these teachers care for their students. They are advocates, having "adopted" their students they watch out for their students' welfare while at the same time challenging students with high expectations, not accepting the "pobrecito" syndrome.

Implications for Professional Training and Credentialing

The preceding analysis has provided an overview of research, policy, and practice as they relate to the education of linguistic minority students of the United States and those educational professionals who also teach them. It is clear that a variety of programmatic efforts have been developed in response to this growing body of students. It has also become evident that professional education training, particularly for teachers, has not kept pace with the demand for specifically trained educational personnel with expertise in these new programmatic endeavors. However, it is not the case that training and credentialing of such individuals has been completely ignored. The following discussion will provide an overview of activities in this domain. Although not exhaustive, the discussion should provide a foundation for understanding the types of issues relevant to training and credentialing a competent linguistic minority teacher. It is appropriate to indicate that other views, some more detailed, are available (see Ada, 1986; Chu & Levy, 1984, 1988; Collier, 1985).

Linguistic Minority Education: An Instructional Innovation

In any discussion of professional training for linguistic minority education, it is important to note that such training is a relatively new enterprise. Not until the mid 1960s did substantial educational initiatives exist in this specialized arena. It was not until 1974 that the U.S. Congress authorized resources for training activities by institutions of higher education in this area of education (August and Garcia, 1988). The recent nature of this innovation, much like similar developments in the field of special education, has spawned many new training programs that are still struggling to establish themselves as legitimate areas of training alongside longer standing programs in elementary and secondary education. This newness is complicated by the nature of the training-program content; that is, this new program just takes a more multidisciplinary perspective. It must be concerned not only with subject matter and pedagogy but also much more directly with language (native language and/or second language) and instruction for populations that are culturally diverse.

The 1980-82 Teachers Language Skills Survey identified the need for 100,000 bilingual teachers if bilingual programs were implemented in schools in which LEP students from one language background were sufficiently concentrated to make such programs feasible. In 1982, there were an estimated 27,000 to 32,000 trained bilingual teachers, leaving 68,000 to 73,000 yet to be trained. Since 168 institutions of higher education graduate approximately 2,000 to 2,600 trained bilingual teachers each year (Blatchford, 1982), the shortage will continue. The Teachers Language Skills Survey reported that, of 103,000 teachers assigned to teach ESL, only 40 percent had received any training in the methods of doing so. It is estimated that at least 350,000 teachers currently need such specialized training (O'Malley, 1981; Waggoner, 1984). Most unfortunate, is the near "study-state" production of language in minority credentialed teachers. In California, for example, a state experiencing record increases in language minority students, the number of teachers credentialed per year in areas related to language minority education, 1982-89, increased by only 5 percent. During this same period, overall yearly teacher credentialing increased by 48 percent (California Commission on Teacher Credentialing, 1990). During this same period there was a general student population increase of 13 percent, but a 45 percent increase in language minority students (Olsen, 1988).

Halcon (1981) and Development Associates (1984) report on the types of training that linguistic minority teachers working in the field have actually experienced. Less than 25 percent of such teachers report graduating from a specific program designed to meet their

408

needs. Instead, most teachers in linguistic minority classrooms have participated in a variety of unsystematic university coursework, district workshops, and federally or state supported in-service training activities. Moreover, the average formal instructional experience of a teacher assigned major instructional responsibilities related to language minority students is less than 3.5 years. Recall that less than 33 percent of instructors in linguistic minority classrooms or in related support roles hold the requisite state credentials (in those states where such credentials are available and in the majority of cases actually mandatory). Such data continue to suggest that linguistic minority education programs are staffed by professionals not directly trained for such programs who might be acquiring their expertise on the job. This situation indicates that the education of language minority students continues to be viewed as a temporary innovation. By their very nature, educational innovations do not have well-developed training strategies or institutional recognition; they must go through a developmental process to achieve the desired goals of status and permanence. Teacher credentialing related to language minority students is still in its "innovation" phase.

Specific Professional Training Issues

On the basis of the foregoing foundation of linguistic minority teacher training, it is proper to consider briefly the actual content of such preparation prior to any discussion of teacher evaluation or credentialing. As with all training endeavors, it has always been incumbent upon the trainers to identify the desired end product of their efforts in some form of performance competencies. The literature abounds with numerous listings of such competencies (Collier, 1985). The most recent and most detailed is presented by Chu and Levy (1988). This list of competencies is derived from a review of federally and non-federally supported linguistic minority training programs presently operating within United States universities. It focuses on some 34 intercultural competencies, no small number, that serve as a foundation for anticipated instructional success of a well-prepared linguistic minority educator. These competencies are organized into knowledge regarding theory, society, and classroom.

The most widely distributed cited list of credential related competencies was developed and published in 1984 by the National Association of State Directors of Teacher Education and Certification. That list, presented in an abbreviated format in Table 2, was a result of combining previous competency lists developed by the Center for Applied Linguistics in 1974 and the Teachers of English to Speakers of Other Languages association in 1975. The list, although not as comprehensive as the Chu and Levy (1988) list, has served as a cornerstone of teacher-training programs and credentialing analysis in the United States. (See Table 1.)

Table 1
NASDTEC Certification Standards*

Content Standards in Bilingual/Multicultural Education (B/M ED)	Possible IHE Course Offerings
1. Proficiency in L1 and L2 for effective teaching	Foreign language and English department courses
2. Knowledge of history and cultures of L1 and L2 speakers	Cross-cultural studies, multicultural education (ME), history and civilization, literature, ethnic studies
3. Historical, philosophical, and legal bases ED and related research	Foundations of BE (or introduction to BE)
4. Organizational models for programs and classrooms in B/M ED	Foundations of BE
5. L2 methods of teaching (including ESL methodology)	Methods of teaching a second language
6. Communication with students, parents, and others in culturally and linguistically different	Cross-cultural studies, school/community relations communities
7. Differences between L1 and L2; language and dialect differences across geographic regions, ethnic groups, social levels	Sociolinguistics, bilingualism

Content Standards in English for Speakers of Other Languages	Possible IHE Course Offerings
1. Nature of language, language varieties, structure of English language morphology	General linguistics; English phonology, and syntax
2. Demonstrated proficiency in spoken and written English	English department courses
3. Demonstrated proficiency in a second language L1 and L2 acquisition process	Foreign language courses Language acquisition
4. L1 and L2 acquisition process	Language acquisition
5. Effects of socio-cultural variables on learning	Language acquisition, ME, cross-cultural studies, sociolinguistics
6. Language assessment, program development, implementation, and evaluation	Language assessment, program development, and evaluation

*These are supplemental standards to the NASDTEC professional education standards required of all teachers.

Recently states and school districts have begun to articulate the actual expected roles and responsibilities of language minority teachers. New Jersey, for example, identifies its expectations in a New Jersey State Board of Education handbook (1991):

Role of Bilingual Teachers

The following responsibilities should be considered by the district when defining the role of bilingual teachers. The bilingual teacher should:

- help identify limited English proficient students;
- participate with administrators in designing a bilingual program that meets the needs of eligible students;
- communicate with ESL and other teachers in planning for the bilingual program students in ESL and special subject areas;
- provide input in areas covered by pupil personnel services;
- apply current research findings regarding the education of children from diverse cultural and linguistic backgrounds;
- develop language proficiency in the native language of the students enrolled in the program and in English;
- have knowledge of techniques, strategies, and materials that aid teaching in two languages;
- structure the use of two languages to systematically make the transition from the native language to English;
- select activities and materials for classroom use which indicate an understanding of the developmental level of the students;
- help students to identify similarities and differences for successful interaction in a cross-cultural setting;
- provide experiences that encourage positive student self-concept; and
- promote and understand the supportive role and responsibilities of parent/guardians and explain the bilingual program to them.

Role of ESL Teachers

The following responsibilities should be considered by the district when defining the role of ESL teachers. The ESL teacher should:

- help identify limited English proficiency students;
- participate with administrators in designing ESL program that meets the needs of eligible students;
- communicate with other teachers in planning for the teaching of the ESL program student in the bilingual or English-only classroom;
- demonstrate awareness of current trends in ESL and bilingual education;
- demonstrate proficiency in English commensurate with the role of a language model;
- use English as the principal medium of instruction in the areas of pronunciation, listening comprehension, speaking, structure, reading, and writing;
- select activities and materials for ESL use which indicate an understanding of the language proficiency level of the students;
- express interest in, and have an understanding for the native culture of the students;
- provide experiences that encourage positive student self-concept; and
- promote and understand the supportive role and responsibilities of parents/guardians and explain the ESL program to them.

Source: Guidelines for Development of Program Plan and Evaluation Summary. Bilingual/ESL Programs and English Language Services, Fiscal Year 1991. New Jersey State Department of Education.

Credentialing and Professional Assessment of Language Minority Teachers

The professional assessment of language minority teachers is a substantially problematic, complex, cumbersome and area "ripe" for criticism. Even more so than the art of teacher assessment in general. It is important to note in this regard that professions are characterized by two broad features (Friedson, 1986): (a) acquisition of knowledge obtained through formal education endeavors, (b) an orientation toward serving needs of the public, with particular emphasis on an ethical and altruistic concern for the client. Therefore, teaching in this country's public schools, and teaching language minority students clearly qualifies as a profession. Given the "professional" nature of this enterprise, a concern for assessment of the professional should not come as a surprise. Assessing professional competence is as old as professionals. According to McGahie (1991), Moses and Jesus Christ set out direct guidelines for assessing religious professionals; Confucius argued that "No man is a good doctor who has never been sick himself"; and, Shakespeare, in the Henry VII soliloquy regarding lawyers, wrote, "Heaven is above all, yet: there sits a Judge, that no king can corrupt." Society or its representatives have been judging the competence of professionals for quite some time. However, it is important to note that like professional themselves, judgments of professional competence are embedded in a local time and place, in line with the professions' "Zeitgeist." That is, these assessments are in concert with the general intellectual and ethical climate and needs of the time (McGahie, 1991).

The assessment of teachers, and language minority teachers is no different. Our present concerns with regard to professional assessment are driven by the ethical considerations of our time and the pressing needs for such professionals. Very specifically, we have relegated the "job" of professional assessment in this country to the states or to professional societies, or, some combination of these institutional representatives. In addition, we have chosen to either focus on assessing the individual as a preprofessional before allowing that individual to enter the profession (usually through examination, the National Teaching Exam is an example), or, we have focused our attention on the assessment of the preprofessional institutions/programs which produce teaching professionals (the NCATE reviews are an examples of "association" reviews while the California Commission on Teaching Credentialing program reviews are examples of state authorized reviews). In some cases, both individual and program review is required.

As is the case for teacher assessment and credentialing of "regular" teachers, the credentialing of language minority teachers is quite variable. Table 2 provides a summary of teacher certification requirements and/or opportunities for specific professional teaching services directed at language minority students. The table identifies the type of teaching credential which are available in all 50 states and U.S. territories along with information regarding that state's or territory's legislative stance regarding such credentialing. These data indicate that 25 states presently do not offer professional credentialing in this domain of the teaching profession. That is, half of the country does not attend to this professional sub-category. These states are not formally interested in any special professional teaching competences related to language minority students. It is not coincidental that those states least impacted by language minority students are those same states which do not address the professional assessment of teachers serving these students. Keep in mind that all states require certification of their public school, teaching professionals.

Table 2
Teaching for Language Minority Students: Evaluating Professional Standards

Teaching Credentials: 1991 State Profiles

State or Territory	Legislation 1*				Teacher Cert 2*	
	MANDATES	PERMITS	PROHIBITS BE	NO STATUTES	BIL EDUCATION	ESL or OTHER
Alabama				✓		
Alaska	✓					
Arizona	✓				✓	✓
Arkansas				✓		
California		✓			✓	✓
Colorado		✓			✓	
Connecticut	✓					
Delaware				✓	✓	
D.C.				✓	✓	✓
Florida				✓	✓	✓
Georgia				✓		
Hawaii				✓		✓
Idaho				✓		
Illinois	✓				✓	✓
Indiana		✓				
Iowa	✓					
Kansas		✓				
Kentucky				✓		✓
Louisiana				✓		✓
Maine		✓				
Maryland				✓		
Massachusetts	✓				✓	✓
Michigan	✓				✓	
Minnesota		✓			✓	✓
Mississippi				✓		✓
Missouri				✓		
Montana				✓		
Nebraska				✓	✓	✓
Nevada				✓		✓
New Hampshire		✓			✓	✓
New Jersey	✓				✓	✓
New Mexico		✓			✓	✓
New York		✓			✓	✓
North Carolina				✓		✓
North Dakota				✓		
Ohio				✓	✓	✓
Oklahoma				✓		
Oregon		✓				
Pennsylvania				✓		
Rhode Island		✓			✓	✓
South Carolina				✓		
South Dakota		✓				
Tennessee				✓		✓
Texas	✓				✓	✓
Utah		✓				
Vermont				✓	✓	
Virginia				✓		✓
Washington	✓				✓	✓
West Virginia			✓			
Wisconsin	✓				✓	✓
Wyoming				✓		
Amer Samoa		✓			✓	
Guam	✓				✓	
N. Marianas				✓		
Puerto Rico				✓		✓
Tr Terr of Pacific				✓		
Virgin Islands		✓				

1* Wether state legislation mandates, permits, or prohibits special educational services for limited-English-proficiency (LEP) students. e.g., transitional bilingual education (TBE). English as a second language (ESL), immersion, and maintenance programs.

2* Wether state offers teaching certification in Bilingual Education, ESL, or other related areas.

Source: U.S. Department of Education, National Clearinghouse for Bilingual Education, (1986). Forum, IX, 3; Updated by each SEA Listed (1991).

Of particular interest is a subset of states which when taken together are home to almost two-thirds of this nation's language minority students: California, Florida, Illinois, New York, New Jersey and Texas. In these states, bilingual credentialing and ESL or some other related credential/endorsement is available. However, in only three of the six states is such credentialing mandated. Therefore, even in states which are highly "impacted" by language minority students, there is no the direct concern for the specific mandating of professional standards. Valencia (1991) has suggested that with the segregation of language minority students, particularly Chicano students in the Southwest, state school systems are not equally affected by these students. Chicano students tend to be concentrated in a few school districts within the state, and even though their academic presence is felt strongly by these individual districts, they do not exert this same pressure statewide. I will return to this important observation, since it identifies a possible alternative forum for professional assessment of significance to enhancing services to language minority students.

Even for those states (a total of 28 states) which address the specific need to assess the professional competence of language minority teachers, the present modes of assessment are highly problematic. unfortunately, the data is quite clear on the problems of individual assessment of teacher professional competence. Present professional assessment can be criticized on several levels (McGahie, 1991; Sternberg and Wagner, 1986; Shimberg, 1983):

1. Professional competence evaluations usually address only a narrow range of practice situations. Professionals engage in very complex planning, development, implementation, problem solving and crisis management. These endeavors do not usually require technical skills and knowledge which are easily measured. The earlier discussion of "effective" language minority teachers (Garcia,1991) exemplifies this complexity.

2. Professional competence evaluations are biased toward assessing formally acquired knowledge, likely due to the preponderance of similar assessment of student academic achievement. We assess teachers like we assess students, even though we have differing expectations regarding these populations.

3. Despite the presumed importance of "practice" skills, professional competence assessments devote little attention to the assessment of enunciated practice skills. With regard to language minority teachers, we do have some understanding of specific skills that "might" be necessary. Although due to the lack of specific research in this domain, I would be hard pressed to articulate the exact skills which I would recommend in need of assessment.

4. Almost no attention is given to what has earlier been identified as the "disposition" and "affective" domains of the language minority teacher. Yet, in recent "effective" teacher analysis, these teacher attributes were identified as significant as content knowledge and practice skills (Pease-Alvarez, Garcia and Espinosa, 1991).

In addition to the above concerns, professional assessment instruments are subject to severe violations of reliability and validity. Feldt and Brennan (1989) have demonstrated that components of measurement error are highly inconsistent in the arena of professional assessment. Similarly, test validity is a fundamental problem for professional assessment (Berk, 1986). Keep in mind that inference about professional competence or ability to practice are actually inferences about specific constructs. This is the old and dangerous "chicken-and-egg-problem." We construct an assessment and soon we are willing to say that whomever scores at "such-and-such" on that assessment is competent. At the base of this assessment however, is the legitimacy of the constructs which generated the assessment. We presently lack any definitive body of research and knowledge regarding the constructs which embody good teachers, in general, and good language minority teachers, specifically. That knowledge base is developing, but it is presently not substantive in nature (Garcia, 1991).

What are we left with? According to McGahie (1991), teacher professional assessment actually is operating within the "connoisseur" model of professional assessment. This model carries certain presuppositions which are relevant to language minority education:

1. Not all features of professional practice can be quantified.

2. There is no "one best answer" to a professional problem or question.

3. Connoisseurs are unbiased, fair in rendering decisions, and due to their demonstrated competence and commitment to the profession and students are the most effective evaluators of teaching professionals.

The connoisseur model is routinely used in a number of professional assessment endeavors like the performing arts and theatre. We would never imagine using a "test" to determine motion picture academy awards. In fact, to determine "Teacher of the Year" honors within local districts, at the state level, and even at the national level, connoisseurs are called upon to serve as judges. They are asked to use their varying experience and expertise to identify the "best." In our own research on "effective" language minority schools, classrooms and teachers, we rely heavily on nominations from connoisseurs -- teachers, administrators, parents and students (Garcia, 1991).

Closer examination of the present mode of teacher training program evaluations indicate that the connoisseur model is the primary model in operations. "Experts" are sent to any program to evaluate the effectiveness of that program. In turn, those local program experts, acting in a connoisseur role evaluate individual teacher candidates.

Is this presently an acceptable model for evaluating language minority teaching professionals? Unfortunately, due to the innovative nature of language minority education -- we are learning how "best" to do it at the same time that we are doing it -- , the limited number of experts/connoisseurs available, and the diversity of students and therefore programs which serve these students, evaluation of language minority teachers is highly problematic. Over time, as we develop a large corp of connoisseurs, it will be possible to utilize this model, and, it is likely the only and best model appropriate. At present, however, it is not possible to implement this model on any large scale with any hope that it will be either reliable or valid.

District Level Credentialing

If the connoisseur model is not possible on a grand scale, it may not be impossible to do well on a smaller scale. Recognizing that the university programs were not, in the short term, able to meet the growing demand for linguistic minority teachers, extensive in-service training initiatives have become the typical vehicle for meeting these growing professional needs. In 1974 federal resources were dedicated to the in-service enterprise, and those resources have continued. Bilingual Education Service Centers conducted needs assessments on a regional basis and implemented regular in-service training activities from 1975 through 1982. In the late 1980s a smaller federally funded effort located in regional Multifunctional Resource Centers continued this activity. In addition, state offices of education in states highly affected by linguistic minority students have developed their own resources for in-service training programs.

Significantly, local school districts have implemented extensive in-service programs to meet their particular needs in substantively increasing the linguistic minority expertise of their teaching personnel. One such program, in Denver, Colorado, exemplifies this in-service training activity. This urban district, highly affected by linguistic minority students, determined that its needs could be partially met by the professional development of its existent teaching staff. Several training presuppositions guided the development and implementation of the in-service training.; (a) teachers needed theoretical grounding and practical application of instruction reflecting that theory, (b) external consultants with linguistic minority expertise would work collaboratively over an extended period of time (4-6 years) with a cadre of local teachers, (c) a local teacher group demonstrating enhanced expertise would provide mentor support to their district colleagues, (d) development of new mentor groups at individual school sites would ensure the systematic augmentation of linguistic minority expertise throughout the district. The district also developed its own "credentialing" requirements, feeling that the state requirements were considerably too generous and left significant holes in requirements. A recent analysis of this in-service strategy indicates that over 500 district teachers participated in this training from the mid 1980s to the late 1980s. Significant gains in service delivery to Denver's growing population of linguistic minority students have been documented. A corp of 100 linguistic minority mentors now exists in support of the over 500 linguistic minority teachers. This mentor corps continues to provide formal training experiences, classroom demonstrations, local site networking, and curricular leadership. These experts or connoisseurs also serve to evaluate new teaching professionals.

What was born out of great necessity in Denver, Colorado, may serve to instruct us regarding the development of language minority teaching professionals and their evaluation. First, professional training takes on a localized characteristic. Such a local emphasis realizes the diversity of students and programs which are present in the local district. Over time, it develops a corp of connoisseurs, and utilizes those locally developed connoisseurs to serve in an evaluative capacity. Therefore, highly relevant local knowledge with regard to language minority education needs is transformed into locally developed experts who in turn evaluate, using local norms, the professional expertise of their colleagues. This is the connoisseur model at its best with regard to the innovative and complex nature of language minority education.

This alternative form of teacher training and district level "credentialing" was born of immediate needs that could not be met through normal teacher training or state level credentialing standards. It demonstrates a useful and highly responsive solution to a problem many school districts face with respect to linguistic minority populations. This alternative form of local training and "credentialing" training could be appropriate for enhancing the effectiveness of most educational professionals, but is worthy of particular attention to the field of language minority education.

Conclusion

It seems clear that language minority students can be served effectively by schools and educational professionals. They can be served by schools organized to develop educational structures and processes that take into consideration both the broader attributes of effective schooling practices and specific attributes relevant to language minority teachers (Carter & Chatfield, 1986; Garcia, 1988; Garcia, 1991; Tikenoff, 1983).

Although the training of language minority education teachers is in a developmental period and in need of further clarifying research, it is clearly not in its infancy. A serious body of literature addressing instructional practices, organization, and their effects is emerging. The training of professional innovators is a challenge for university and federal, state, and local educational agencies. The needs are great, and the production of competent professionals has lagged. However, professional organizations, credentialing bodies, and universities have responded with competencies, guidelines, and professional evaluation tools. These evaluation tools are problematic with regard to their reliability and validity. The most often utilized professional evaluation model is the "connoisseur" model At the state level, this model is problematic. However, local school districts have also had to engage in substantial training endeavors and they have or can develop professional evaluation models, locally derived credentials, with locally developed connoisseurs. This alternative, district level credentialing process is worthy of serious consideration. The challenge for all those engaged in such an enterprise is to consider the rapidly expanding literature regarding linguistic minority teachers, to evaluate its implications critically and to apply it to local language minority education contexts, with a dependency on locally developed connoisseurs.

References

Ada, A. F. (1986). Creative education for bilingual teachers. Harvard Educational Review, 56(4), 386-394.

August, D., & Garcia, E. (1988). Language minority education in the U.S.: Research policy and practice. Springfield, IL: Charles C. Thomas.

Baker, K., & deKanter, A. A. (1983). An answer from research on bilingual education. American Education, 19(6), 40-48.

Berk, R. A. (Ed.). (1986). Performance assessment: Methods and applications. Baltimore, MD: John Hopkins University Press.

Carter, T.P., & Chatfield, M. L. (1986). Effective bilingual schools: Implications for policy and practice. American Journal of Education, 95(1), 200-234.

Castaneda v. Pickard. 648 F. 2d 989, 1007 (5th Cir. court, 1981).

Chamot, A., & O'Malley, J.M. (1986). A cognitive academic learning approach: An ESL content based curriculum. Wheaton, MD: National Clearinghouse for Bilingual Education.

Chu, H., & Levy, J. (1988). Multicultural skills for bilingual teachers. NABE Journal, 12(2), 17-36.

Collier, J. P. (1985). University models for ESL and bilingual teacher training. In National Clearinghouse for Bilingual Education (Ed.), Issues in English language development (pp. 81-90). Rosslyn, VA: Editor.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. Review of Educational Research, 19, 222-251.

Development Associates. (1984, December). Final report: Descriptive study phase of the national longitudinal evaluation of the effectiveness of services for language minority limited English proficient students, Arlington, VA: Author.

Feldt, L. S., & Brennan, R. C. (1989). Reliability. In R. L. Linn (Ed.), Educational Measurement (3rd ed., pp. 105-146), New York, NY: American Council on Education & Macmillan.

Friedson, E. (1986). Professional powers. Chicago, IL: University of Chicago Press.

Garcia, E. (1983). Bilingualism in early childhood. Albuquerque, NM: University of New Mexico Press.Garcia, E. (1988). Effective schooling for language minority students. In National Clearinghouse for Bilingual Education (Ed.) New Focus. Arlington, VA: Editor.

Garcia, E. (in press). Hispanic children: Theoretical, empirical and related policy issues. Educational Psychology Review.

Garcia, E. (in press). Attributes of effective language minority teachers: An empirical study. Journal of Education.

Garcia, E. (in press). Characteristics of effective teachers for language minority students: A review (Working paper #1). National Center for Research on Cultural Diversity and Second Language Learning.

General Accounting Office. (1987). Research evidence on bilingual education (GAO/PEMD-87-12BR). Washington, DC: Author.

Hakuta, K. (1985). Mirror of language: The debate on bilingualism. New York: Basic Books.

Hakuta, K., & Garcia, E. (1989). Bilingualism and education. American Psychologist, 44(2), pp. 374-379.

Hakuta, K., & Gould, L. J. (1987). Synthesis of research on bilingual education. Educational Leadership, 44(6), 38-45.

Halcon, J. (1981). Features of federal bilingual programs. NABE Journal, 6(1), 27-39.

Hudelson, S. (1987). The role of native language literacy in the education of language minority children. Language Arts, 64(8), 827-841.

Hammerly, H. (1985). An integrated theory of language teaching, Burnby, Canada: Second Language Publications.

Krashen, S. D. (1982). Principles and practices in second language acquisition. Oxford, England: Pergamon Press.

Krashen, S. D. & Biber, D. (1988). On course: Bilingual education success in California, Sacramento, CA: California Association for Bilingual Education.

Lau v. Nichols, 414 U.S. 563 (1974).

McGahie, W. C. (1991). Professional competence evaluation. Educational Researcher, 20(1), 3-9.

McLaughlin, B. (1985). Second language acquisition in childhood: Vol. 2. School-age children. Hillsdale, NJ: Lawrence Erlbaum.

National Association of State Directors of Teacher Education and Certification (1984). NASDTEC Certification Standards. Washington, DC: Author.

O'Malley, M.J. (1981). Children's and services study: Language minority children with limited English proficiency in the United States. Rosslyn, VA: National Clearinghouse for Bilingual Education.

Ovando, C. J., & Collier, V.P. (1985). Bilingual and ESL classrooms: Teaching in multicultural contexts. New York: McGraw-Hill.

Paulston, C. B. (1980). *Bilingual education: Theories and issues*. Rowley, MA: Newbury House.

Pease-Alvarez, C., Garcia, E., & Espinoza, P. (in press). Effective instruction for language minority students: An early childhood case study. *Early childhood Research Quarterly*.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: a review. *Elementary School Journal*, *83*(4), 426-452.

Ramirez, A. (1986). *Bilingualism through schooling*. Albany, NY: State University of New York Press.

Reynolds, A., & Elias, P. (1991). *What is good teaching: A review of the literature*. Princeton, NJ: Educational Testing Service.

Rossell, C., & Ross, J. M. (1986). *The social science evidence o bilingual education*. Boston: Boston University.

Schneider, S. G. (1976). *Revolution, reaction or reform: The 1974 Bilingual Education Act*. New York: Las Americas.

Schumann, J. H. (1976). Affective factors and the problem of age in second language acquisition. *Language Learning*, *25*, 209-239.

Seliger, H. W. (1977). Does practice make perfect? A study of interactional patterns and L2 competence. *Language Learning*, *27*(2), 263-278.

Seliger, H. W. (1984). Processing universals in second language acquisition. In F. R. Eckman, L. H. Bell, & D. Nelson (Eds.), *Universals of second language acquisition* (pp. 161-183). Rowley, MA: Newbury House.

Shinberg, B. (1983). What is competence? How can it be assessed? In M. R. Stern (Ed.), *Power and conflict in continuing professional education*. Belmont, CA: Wadsworth, pp. 17-37.

Sternberg, R. J., & Wagner, R. K. (Eds.). (1986). *Practical intelligence*. New York, NY: Cambridge University Press.

Tikenoff, W. J. (1983). *Compatibility of the SBIF features with other research instruction of LEP students* (SBIF-83-4.8/10). San Francisco: Far West Laboratory.

Troike, R. C. (1981). Synthesis of research in bilingual education. *Educational Leadership*, *38*(6), 498-504.

Equal Education Opportunities and Transportation of students Act of 1974, §294(f), 20 U.S.C.

U.S. v. Texas, 647 F. 2d 69 (9th Cir. 1981).

Valencia, R. (1991). Chicano school failure and success. New York, NY: The Falmer Press.

Villegas, A. M. (1991). Culturally responsive pedagogy for the 1990's and beyond. Princeton, NJ: Educational Testing Service.

Waggoner, D. (1984). The need for bilingual education: Estimates from the 1980 census. NABE Journal, 8(2), 1-14.

Willig, A. C. (1985). A meta-analysis of selected studies on effectiveness of bilingual education. Review of Educational Research, 55, 269-317.

Wong-Fillmore, L. (1976). The second time around: Cognitive and social strategies in second language acquisition. Unpublished doctoral dissertation, Stanford University, Palo Alto, CA.

Wong-Fillmore, L. & Valdez, C. (1986). Teaching bilingual learners. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed. pp. 648-685). New York: Macmillan.

Evaluating Limited English Proficient (LEP) Teacher Training and In-Service Programs

Stephanie Dalton
National Center for the Study of Cultural Diversity
and Second Language Learning

and

Ellen Moir
Board of Studies in Education
University of California, Santa Cruz

The topic of this paper, Evaluating LEP Teacher Training and In-Service Programs, is, unfortunately, among the least reported issues in the literature of teacher education research. The consequences of this neglect are evident not only in program evaluation's underdevelopment and in unexamined teacher education programs but also in the individual experiences of increasing numbers of teachers nationwide. LEP Teacher training and in-service programs can provide teachers with the assistance necessary to increase the academic performance of linguistically and culturally diverse students. There is a growing knowledge base regarding the content and pedagogy of effective education of LEP students. In addition, qualitative and quantitative methodologies offer diverse measures for obtaining useful data about program experience to apply in program development.

In this paper, the authors first briefly review the history of teacher education evaluation, particularly its methodology, and then examine content recommendations coming from current research on effective education of linguistically diverse students. Secondly, they report their experiences with an evaluated preservice and an in-service teacher education program. The preservice program was a University of Hawaii alternative program titled Preservice Education for Teachers of Minorities (PETOM) and the in-service program was part of the California New Teacher Project (CNTP) at the University of California at Santa Cruz. Based on the presentation of teacher education program evaluation literature, the findings of recent research on effective teaching and learning models for linguistic minorities, and the experiences of the preservice and in-service programs, the paper will conclude with recommendations for LEP preservice and in-service teacher education program evaluation. The recommendations place emphasis on generating evaluation designs that yield useful information for teacher education program developers and faculty, the use of broad based methods to obtain multiple perspectives on program experience and effect, and the desirability of program evaluation design that informs and involves all participants.

Introduction

A literature search in 1988 for documentation of multicultural program evaluation produced a single reference. This situation has changed little since for programs addressing educational issues of cultural or linguistic diversity. To underscore the void in this domain, teacher education program evaluation, in general, has been referred to as "teacher education's orphan" (Galluzzo and Craig, 1990).

Cooper's (1983) summary of program evaluation in teacher education included a very short list of institutions (such as the University of Georgia and Ohio State University) where evaluation was an integral part of teacher education program development. Although few examples exist in the literature, survey responses indicate that more evaluation is actually practiced. In the same year as Cooper's summary, Adams & Craig (1983) conducted a survey of institutions affiliated with the American Association of Colleges for Teacher Education (AACTE). The survey, a follow-up questionnaire, was mailed to teacher education programs across the nation. The survey identified four hundred institutions that gathered program evaluation data.

These reports suggest a possible mismatch of program evaluation literature and practice. It appears that program evaluation is indeed conducted but primarily for accountability reports to accrediting or other external agencies. Certainly program evaluation design must meet the reporting requirements and criteria of official agencies such as the National Council for Accreditation of Teacher Education (NCATE), State Commissions for Teacher Credentialing and the National Association of State Directors for Teacher Education Certification (NASDTEC).

It appears that in practice little program evaluation is specifically designed for internal use in program improvement or to increase understanding about developmental processes. This means that the suitability of teacher education curricula for the communities served, the effect of program on professional and LEP student consumers, and experiences of program participants remain largely unexplored. With so few examples of evaluated programs available, teacher education programs experience little pressure to evaluate. Evaluation's low priority in preservice and in-service teacher education program development is one explanation for programs' unresponsiveness to rapidly changing teaching conditions.

First Recommendation

Our first recommendation is that LEP teacher education and inservice programs employing evaluation need to be identified and documented. More examples will provide the models, explore the process, and stimulate the installation of an evaluation component in programs.

We know the nation is facing a major challenge in reshaping its schools to be appropriate to the diversity of its population. The public school demographic trends are changing for teachers and students. For example, the state of California has positions for 11,000 bilingual teachers. Last year, the University of California trained 272 bilingual teachers. An OBEMLA sponsored forum on "Staffing the Multilingually Impacted Schools of the 1990s" produced evidence from the participating administrators indicating a current shortage of 175,000 bilingual teachers nationwide if a 20:1 student-teacher ratio is considered. Demographic imbalance between non-minority teachers and minority students means that teachers will be working with students whose backgrounds are culturally and linguistically different from their own. Sometimes, as in the case in the Los Angeles School District, veteran teachers are finding the demographics of their classroom changing from year to year and from familiar cultural and linguistic backgrounds to those that are totally unfamiliar.

LEP teacher training and in-service programs can provide teachers with the assistance necessary to increase the academic performance of linguistically and culturally diverse students. While there is yet no template for effective LEP student instruction, there is a substantial knowledge base regarding the content and pedagogy of effective education of LEP students. For example, the National Center for Research on Cultural Diversity and Second Language Learning is currently engaged in identifying and documenting the instructional practices of effective teachers of culturally and linguistically diverse minorities. These findings will provide the working models of effective instructional practice informing the continuing debate regarding the education of linguistic minorities.

Teacher education programs can use these effective models for educating LEP students as resources for their own program development. Programs desiring to import information from models, the knowledge base on effective education for linguistic and culturally diverse students, new research findings, and other sources design evaluation to aid this process. The University of Hawaii and California programs described in this paper have program evaluation components. The notion of evaluation design emerging from these program experiences is one that obtains useful feedback about program operations as well as processes. Importantly, in this notion is the means that programs use to respond to evaluative feedback and discover their own developmental processes.

Second Recommendation

The second recommendation is directed to the issue of program responsiveness. Program responsiveness to evaluation feedback needs documentation. Evaluation methodology designed to yield relevant, substantive, and useful information for program developers and faculty is most likely to result in program responsiveness to feedback. More examples from evaluated programs will encourage this practice and demonstrate useful evaluation design.

The history of the documentation of teacher education evaluation, particularly its methodology, is worth reviewing for models of effective and less effective LEP program evaluation documentation. Content recommendations coming from current research on effective education of linguistically diverse students relate to evaluation issues and influence program evaluation form and construction. The trends and themes in the history of teacher education program evaluation can inform LEP teacher education program evaluation.

Although there is no evidence that linguistically diverse students were ever norm group members or that models influencing evaluation design were validated on them, the record of experience is useful for understanding what methods produce useful information for program developers. The earliest evaluation reports dating from the 1940s relied on models of evaluation emphasizing goal attainment, product orientation, and teacher performance competency based on program objectives.

A 1944 publication by Troyer and Pace described methods used by teacher education institutions to evaluate programs. They identified program components such as the general education component, the professional education sequence, student teaching and follow-up studies and described a variety of methods for assessing the skills, attitudes, and understandings of preservice teachers. According to Galluzzo and Craig (1990), teacher education evaluation has only minimally shifted since the 1944 documentation. As in an AACTE survey by Adams & Craig (1983), program evaluation reports continue to rely heavily on single tap follow-up surveys.

The preponderance of follow-up studies that consist of postgraduates' self-reports raise issues of reliability and validity. The usefulness of such data for program development has been rightly challenged. For example, Katz et al. (1981) critically reviewed 26 studies using graduates' responses to follow-up questionnaires as evaluation data. Here Katz et al. contributed the "feed-forward" principle. Feed-forward is defined as the "resistance from the student at the time of exposure to given learnings and, later, protestations that the same learnings had not been provided, should have been provided or should have been provided in stronger doses" (p. 21). This situation illustrates the substantial validity limitations on data gathered through questionnaires from individuals after the experience of the program. Katz et al. call for evaluation data that will be more informative for program effectiveness and development.

In 1970, Sandefur's monograph on a model for program evaluation described a product-oriented, competency-based approach. In this outcomes based model, selected competencies served as objectives of the program for student and new teacher performance. This model has been re-energized in the 1987 National Council for Accreditation of Teacher Education (NCATE) guidelines which emphasize the degree to which students achieve objectives of a program. Medley (1977) in addressing teacher education program evaluation defined it as the extent to which "the training experiences produce the competencies defined as objectives of the training program" (Medley, 1977, p.69).

In the 1980s, the value of alternative approaches including increased description and documentation of on-going program experience from the participants' point of view were recognized and called for. Qualitative methodology was recommended for capturing program features previously ignored. These features included program antecedents, contexts of program operation, intended audience, developmental and process issues. This approach meant previously unexplored issues relating to student and teacher language and thinking development could be studied.

In sum, design and implementation of on-going teacher education program evaluation is currently defined in wider ranging terms than the narrowly conceived outcomes-based orientation that has been predominant. Methodology that has been overladen with self-report particularly from graduate surveys has produced data that program developers are hard pressed to find a use for. The link between actual program practices and teacher development, as a result, is substantially unexamined.

Third Recommendation

Our third recommendation for program evaluation supports methodological approaches employing quantitative and qualitative inquiry methods for the purpose of exploring processes and transformations produced by programs and experienced by preservice and inservice teachers. The program itself is engaging a process in designing and conducting evaluation and this is worthy of attention. The use of broad based methods and measures that obtain multiple perspectives on program experience and effect is most likely to capture information about developmental processes and the links between program and sources of influence.

Issues of LEP Student Education

In reviewing the recommendations coming from current research about augmenting linguistic minority student achievement, consensus on every issue is rare, but understanding of the issues in the form of a substantial knowledge base is accumulating. New data are providing better understanding of the cognitive assets associated with high degrees of bilingualism (Díaz, 1983, Hakuta, 1986). Language learning research is indicating that second language development is not significantly impeded by native language and human cognition is indeed organized to accommodate new language learning (McLaughlin, 1990). The position that bilingualism is a deficit condition is no longer sustainable.

Among these themes surrounding the education of LEP students, one of the most central concerns is the instructional use of the languages of bilinguals in classrooms (Garcia, 1990). Recent research reports of Wong-Fillmore, Ramirez and the growing influence of Vygotskian theory inform our understanding about the critical role of language in the education of LEP students.

Wong-Fillmore's (1991) position that children of linguistic minorities are assimilating into English at the expense of their home language is another theme of the debate. Capitalizing on the common belief that the younger a learner is the faster and more completely a new language can be learned, states have legislated younger and younger English-only instruction. What has been ignored in this policy is the cost to the young learners and their families in primary language loss -- what has been referred to as "subtractive bilingualism" (Wong-Fillmore, p. 1, 1991). New language learning is not dependent on inattention to native language and the social consequences of this approach to LEP student education have been made sadly clear.

Ramirez' (1991) longitudinal study compared the relative effectiveness of three bilingual programs: (1) structured English immersion strategy, (2) late-exit and (3) early-exit transitional bilingual education. The study's findings strongly support the effectiveness of bilingual programs and indicate that students exposed to more English in English Immersion programs perform no better overall on tests of English language ability than do students in early- or late-exit bilingual programs.

Of particular interest for this topic, Ramirez found teachers' responses to the challenge of heterogeneous LEP classrooms to be remarkably similar in that a minimum of language production opportunities were provided for the students. When students were separated by language classifications of LEP, Fluent English Speakers (FEP) and English Only (EO), there was no difference in teachers' talk among the three programs. In mixed classes of LEP, FEP and EOs, teachers' talk did differ in that they explained, modelled and monitored more often asking fewer questions, giving fewer instructions and less feedback. In other words, teachers talked to the students more while asking the students for less talk. For all students, there was scant dialogue or instructional conversation. Teacher talk predominated in all the programs at two times the rate of student talk. When teachers and students did interact students' responses were frequently nonverbal or simple information recall statements.

This evidence presented in the Ramirez study supports the persistence of the assessment or recitation model of instruction which minimizes social interaction and student language production. The "script" of the recitation model consists of assigning text material to students, asking them to "recite" from it, most often through quiz, worksheet, or test, and assessing whether or not the students learned it. This teaching-by-assessment is in contrast to teaching-by-assistance, a model of teaching associated with the theoretical work of L.S. Vygotsky (Tharp & Gallimore, 1988).

From Vygotsky and his disciples and elaborators, we are coming to know how teaching takes place in ordinary interactions of everyday life and results in the generation of higher order thinking. "The developmental level of a child is identified by what the child can do alone. What the child can do with the assistance of another defines what (Vygotsky) called the zone of proximal development....It is in the proximal zone that teaching may be defined." In Vygotskian terms, teaching is good only when it 'awakens and rouses to life those functions which are in a stage of maturing, which lie in the zone of proximal development.' (Vygotsky, 1956, p. 278, quoted in Wertsch & Stone, 1985, Tharp & Gallimore, 1988, p. 4). In this redefinition of teaching, teachers assist students through their Zones of Proximal Development anticipating, selecting and maximizing the moments to assist student performance.

Of the many ways to assist performance, dialogue, the ability "to form, express, and exchange ideas in speech and writing" is fundamental to the development of thinking skills and "is the way parents teach their children language and letters." Dialogue or, in Tharp & Gallimore's term, the Instructional Conversation, occurs through "the questioning and sharing of ideas and knowledge that happens in conversation" (Tharp & Gallimore, 1988, p. 5). The Ramirez finding that teachers in the study generated few language production opportunities for LEP students raises questions about teachers' knowledge base and skill level in regard to language development and its relationship to cognition. In a Vygotskian conception of teaching and learning, the absence of dialogüe or instructional conversation seriously constrains development of students' higher level thinking and complex learning.

Fourth Recommendation

Therefore our fourth recommendation for LEP teacher education program evaluation is that it identify and examine programmatic features and teachers' experiences important to language development. Language development features include teacher student ratio of talk, progression and level of student teacher talk, and student opportunities for talk using school language and students' first language or dialect. Data collection strategies for language development include observation, informal records and student interactive journal entries, and other productions. Alternate data collection strategies for other programmatic features include descriptive summaries and samples of student materials.

Examples of Preservice and In-Service Program Evaluation

The program evaluation experience of two programs, a University of Hawaii alternative program, Preservice Education for Teachers of Minorities (PETOM), and the California New Teacher Project (CNTP), at the University of California at Santa Cruz (SCCNTP), provides examples of veteran and preservice teacher education program evaluation.

PETOM Preservice Program Experience

In a state with a majority population of linguistic minorities such as Samoan, Tongan, Filipino, Laotian and Hawaiian Creole or pidgin dialect speakers, many students', and particularly those at-risk, sole experience with Standard American English occurs only in the school setting. The Native Hawaiian Educational Assessment Project described "Persons of Native Hawaiian ancestry" as those who "have suffered disproportionately from educational and social inequality for some time. Descendants of the original inhabitants of Hawaii find themselves at the bottom of indicators of success in modern America, and they are sometimes referred to as "strangers in their own land" (1983, p. 3). As a teacher educator and alternative teacher education program developed at the University of Hawaii, oral and written language development of young people was as central a curricular issue for teacher preparation as multicultural understanding and sensitivity.

431

The faculty of an experimental teacher education program, titled Preservice Education for Teachers of Minorities (PETOM), was committed to preparing teachers to provide at-risk minority students with experiences in language leading to the attainment of literacy and the ability to function in mainly verbal settings such as those of classrooms and, eventually, the workplace, community and society.

This commitment was underscored by the inclusion of a semester course in Language Development within the program's two-year, field based curriculum. PETOM's efforts to evaluate its program quality and student development were collaborative producing and evaluation report each year for six years. Three features of PETOM, the program's origins, curriculum, organization and conceptual framework relate to the program's evaluation experience.

First, PETOM grew from the work of the Kamehameha Early Education Program (KEEP) teacher consulting model. In KEEP's model of in-service teacher consultation, evaluation was continuous taking the form of weekly observation and feedback collegially provided by a peer consultant and quarterly data feedback provided by criterion-referenced testing. KEEP's in-service teacher training effectiveness was evaluated by the use of standardized test scores of Hawaiian, part-Hawaiian students' reading skills. While undoubtedly successful, KEEP's in-service teacher training program invested 3 to 5 years in teacher development.

PETOM, as an outgrowth of KEEP, sought to capitalize on the preparation, practice, and reflection time available to preservice teachers during their professional training. A preservice teacher education program incorporating the principles of KEEP combined with extensive field experience in classrooms serving diverse students could facilitate preservice teachers' developmental course from novice to skilled.

Secondly, PETOM's curriculum included generic methods emphasizing integration of the content areas, language development, inquiry process, classroom management, child development and foundations. Faculty objectives for the course work included the translation of theory into practice and maximal modeling of instructional strategies. Course work built directly on the novice teachers' field experiences.

Thirdly, PETOM's faculty worked collaboratively to develop and implement the program. The Methods Instructors commented about their experience saying,

> "Traditional instruction for preservice teachers is scattered among different faculty members who rarely talk with one an-

> other. No single group is responsible for seeing the preservice teachers through their entire program. As Eisner pointed out, university faculties may be even more isolated than elementary classroom teachers. Teacher educators rarely know what the other is teaching and rarely, if ever, observe one another teach. PETOM has provided an opportunity to break down this isolation" (Picard & Young, 1990, p.31).

The program's theoretical framework, a Vygotskian conception, sensitized the faculty to the ways that teachers assist students to learn and teacher educators assist preservice teachers. The faculty struggled to understand the development of teachers' professional thinking and skills by the Vygotskian principles; as students learn in conversation with their teachers, teachers learn in conversation with their faculties and in the classrooms of their early field experiences. PETOM's preservice teachers were assisted through their proximal zones, by their instructors and/or peers in program activity including classroom interactions, course discussion, interactive journals and peer and/or instructor review of lesson videotapes.

Again PETOM Instructors described it saying,

> "there are times when we may not seem to be clear as to whether we are talking about the education of teachers or the education of children. The truth is that many of the principles that follow apply to both. Our expectation is that teachers will teach as they have learned" (Picard & Young, 1990, p. 31).

As an experimental program within the College of Education, PETOM conducted annual program evaluations. Over the course of three cohorts of two years duration each, several evaluation designs were used. The findings were reported to the Dean, relevant College of Education committees, and other audiences. However, of the greatest importance was the feedback to the faculty for program development purposes. The faculty was responsive to positive and critical feedback provided directly by the students, revising course work requirements and on communication issues. All of the evaluation methodology used represented faculty interest in obtaining feedback about developing their own and teacher performance competence in conceptual and specific skill areas emphasized in the program.

The measures used and listed in Appendix A included checklists of the College of Education for rating performance competence of the PETOM teachers. PETOM checklist data was compared to control group checklist data from the traditional program. Demographic data, such as ethnicity, age, grade point average and place of birth, were compiled demonstrating PETOM's commitment to diversity in its recruitment and retention. PETOM students and graduates were interviewed about their program experience and asked to rate their program experiences and the program components. Cooperating Teachers were interviewed about their experiences in the program. Principals in the schools where PETOM students were hired as first year teachers were interviewed for their perspective on the PETOM teachers performance. Situational probes were devised to discriminate between PETOM students and the control group for competence in program emphases such as language development and the use of culturally compatible teaching strategies. Open ended, stimulated recall interviews of the preservice teachers were a requirement of the field experience. Some of these were conducted and coded. Of these methods, the students' ratings, principals' comments, situational probes, and open-ended interviews will be presented and discussed.

Student Ratings

After the first cohort of PETOM students had taught, they were given a survey questionnaire to rate the usefulness and relevance of their learning experiences in the program to their subsequent teaching. Their ratings demonstrated that the students considered their general undergraduate course work (in their case, the first two years of university general education requirements) barely relevant (or 2 out of a possible 5 points) to their teaching preparation. In contrast, the field experience opportunity was considered most relevant with a rating of 4.5 out of a possible 5 points. The education course work was also very positively valued at just under 4.5.

Immediately after the third cohort of students completed the program, they were asked to rate questions about their PETOM education. In the ratings, there is a positive trend in general. However, the pattern emerging from the rank ordering is one that indicates the graduates favored the more participatory kinds of experiences such as student teaching and observation-participation over the course work and the field experience seminars (Speidel, 1990, p. 77).

While it is desirable to understand the perspectives of program graduates, the value of this type of feedback to the faculty and program is limited. These measures validate novice hunger for hands-on experience and give little information about teacher development. To obtain more useful data for their purposes, the faculty decided to obtain data from the principals hiring the program graduates.

Principals' Comments

Overall the Principal's comments were positive describing the teachers as good, fine, excellent, and strong. The principals' negative comments included such statements as: (1) the teachers had difficulties applying their knowledge due to lack of experience (3 times) and (2) the teachers need more relevant field experience (3 times). One principal commented on a teacher's unsmiling demeanor. In the narratives, principals described teachers indicating their own criteria for good teaching. One example is:

Principals' Narrative Data Samples:

> "I think Bev is a really good example of an excellent teacher. She is just beginning now to take some responsibility for the rest of the faculty in sharing some of her ideas. She attended a workshop where she had to come back and share what she had learned with the rest of the faculty and she did an excellent job. I look upon her as someone who will someday become Teacher of the Year. She is unassuming and sincere in her efforts. Bev is always open and eager to learn. I think either your program did a good job on her or else she is just a natural teacher. I think your program should take some of the credit. Beverly is exemplary."
>
> Joan and the other two PETOM graduates now teaching at my school are very enthusiastic and always willing to learn. They have a good knowledge base and are well equipped with effective teaching strategies although being new teachers they are still having difficulties applying them.
>
> They are middle-class teachers dealing with low-income, severely at-risk youngsters and this is an incredibly hard task. Therefore, they have had to make a lot of adjustment this first year and in many ways they are not yet fully equipped. I look at these teachers as slowly evolving and in two more years they will be top-notch. But they are never frustrated or depressed. They are lovely people who are trying hard and maintaining very good attitudes despite the many obstacles that working with deprived youngsters bring.
>
> Working on the Leeward Coast is always a very frustrating experience for very new teachers because most of them have not had the hands-on time with these kinds of kids. They have to first experiment with and weed through a multitude of teaching strategies before finding those that are most effective. These strategies must address all kinds of kids and not just those found at Kamehameha Schools or in town. Most young teachers have not had enough field experience in this area and it would really help if they spent time out here before graduating from school....I think your program is wonderful but your graduates need to spend more time in areas such as this before they can be truly competent new teachers. These teachers are growing this year and they are always willing to learn. I am sure they will evolve into first-rate teachers."

For the faculty, principals' comments were validating in that the PETOM students were generally successful as beginning professionals in a school setting. This data was informative about contexts of school culture and community and revealed the principal's belief in the value of "hands-on experience" and the artistry of teaching. The data were less useful for increasing faculty understanding about the effectiveness of their course work. The teacher characteristics highlighted in these comments, while important, were only generally related to course work objectives.

The faculty decided to design measures for collecting course work specific data from PETOM preservice teachers. Situational probes were constructed by the instructors for each course completed during the year of the evaluation. The written probes explored for the effect of the experience of PETOM on its preservice teachers by asking them to provide substantive responses when presented with written situations. It was hypothesized that PETOM student responses would be program specific and substantively richer than those of a control group of students from the traditional program. An example of a Situational Probes developed by the Language Development course Instructor follows:

Situational Probe

> You've organized your second grade class into small instructional groups, but you've noticed that the discussions you've been having in these groups haven't been going as well as you'd like. The students typically give one word overly brief responses. You'd like them to give lengthier answer and elaborate on each other's responses.

a. What will you do in an effort to improve the discussions?

b. What are your reasons for suggesting these actions?

Sample answers given by PETOM students:

1. Ask open-ended questions, questions that allow for more divergent responses. Ask student to elaborate answers. Open/divergent questions require more than one word responses and also give teacher and students a "forum" for extending answers -- ask students to explain response, add to it.

2. I would try to state questions in such a way that requires more than one word answers. Questions could begin with "Why do you think...?," "How...?," "What could/would happen if...?," "What do you think about...?"...However, I think that all this must be taught to the students; they need to be guided through a discus-

436

sion. At first, the teacher must take the lead but the ultimate goal would be for the teacher to only help focus and clarify in a discussion.

3. I would draw on the children's background and experiences. By drawing on experiences, the discussion becomes more interesting and students will feel comfortable in contributing to the discussion.

Sample Non-PETOM Responses:

1. Ask the students to answer in a sentence form.

2. Make rules regarding the discussions: Everyone needs to respond with at least three sentences and respond to someone else's contribution to the discussion. By providing rules to the discussion and requiring each child to participate this would help to increase the length of the answers and elaboration.

3. To improve the discussions I would try having one large group in which students raise their hands to respond to questions or answers. The reason for suggesting this is that this particular group may not work well in groups. As a group (large one), the students may respond better instead of in smaller groupings.

The data collected with the situational probes distinguished the PETOM students from non-PETOM students. This was not the case with every probe but, in general, the data indicated that the students were appropriating conceptual material featured in the program's course work. However, the faculty noted that this paper and pencil task did not reveal the students' reactions during the acts of teaching. The latter appeared to be the most critical information about the program effects and the most promising for understanding the teacher development process. In the interviews to be discussed next, the faculty found those data, and they were revealing indeed.

Interview Data

Part of PETOM's field experience required that the preservice teachers videotape themselves teaching. In interviews about their videotaped lessons, the preservice teachers began to disclose their progress through a rigorous developmental process of appropriating and applying course work concepts.

A preservice teacher describes her experience:

"Sometimes, too, when you're told things and when you actually do it, it comes out different. That's what I'm finding out, too.

> They cannot always be done like how (you're told in courses) or there isn't any one best way to do something. Sometimes, too, it just depends on the situation and what's going on. It's really hard."

Although she arranged her teaching activities according to course work guidelines and recommendations, the preservice teacher found that the "situation" or the social conditions of teaching compounded the challenge. Such occasions of purposeful social activity, in Vygotsky's conception, are the basis for forming and transforming thinking. Such data from the perspective of the participant reflects the process of individual development and has the potential to inform program development and expand the knowledge base.

Another preservice teacher said:

> "I don't think I really thought too much about the kind of preparation and the time it would take to do research and to absorb it and understand it enough to teach it to someone, to break it down into steps. That, I kind of had a hard time adjusting to the students' need, and how you can better move them to the point, more guided."

This statement describes the challenge that "guided" or teaching that assisted performance makes on the teacher. The preservice teacher in using terms like "adjusting to the students' need" and "how you can better move them to the point" is operationalizing her understanding of the zone of proximal development. She considers the amount of assistance the student requires based on what she believes he can do with assistance that he cannot do independently. In her words, she will "break it down into steps" in order to assist the student to the "point" or lesson goal.

Typically, the complexity of sophisticated interaction between teacher and students is camouflaged from the novice. It is in the activity itself that the challenge of assisting students to learning goals becomes crystal clear to the preservice teacher. Tharp and Gallimore emphasize that "Assisting performance through conversation requires a quite deliberate and self-controlled agenda in the mind of the teacher, who has specific curricular, cognitive, and conceptual goals" (Tharp & Gallimore, 1988, p.5). Measures that collect data about participants' experience from their point of view during the activity of teaching produce data of use to faculties for understanding course effect and the process of teacher development.

Teacher education and professional development programs emphasizing performance assistance for linguistically and culturally diverse students, heighten preservice teachers' sensitivity and respect for their students' linguistic and cultural differences. LEP programs'

438

aim to assist the performance of preservice and veteran teachers by equipping teachers with appropriate strategies and techniques for effectively educating diverse students. These strategies are designed to remove many of the barriers to communication by adjusting the dynamics of teacher student interaction to be compatible with students' preferences.

Teachers' application of such strategies can be assisted through observation and feedback as well as interview. In the following interview excerpt, a preservice teacher describes her use of one strategy for discussion labelled, "talkstory" and she also describes her feelings. She says:

> "The talk story format...seems to be more comfortable, share whatever they're thinking at the moment...I like it. I don't mind it in a small group. I think I can handle it better but I know when it gets bigger, 5 is very good for me and if it gets to like 10, it gets harder to manage and I have a harder time hearing. I have to cue a lot more about I can't hear so and so talk. Somebody else is talking at the same time I have to keep cueing and reminding them about it. I can only hear so many people at a time. But I feel pretty comfortable with it. I enjoy it and I know the kids enjoy it. So it's a lot more comfortable and natural.... cause they seem to participate more. They get more excited about it and motivated."

The information from the interview data included information about teacher affect, teacher attitude toward the students, and the teacher's progress appropriating the strategy. Her discussion also revealed her use of another technique, cueing, which she was applying simultaneously with talk-story. Her facility with the "talk story" technique and her comfort level in using it indicates her progress from novice to skilled professional. This is valuable feedback for faculty and program developers.

The interviews provide rich data suggesting linkages and interactions among the program, the knowledge base, and the teacher development process. Analysis and interpretation of this data within the context of the total program evaluation holds promise for informing program development. In-depth interview data promises to reveal not only more about processes of teacher development but more about program development processes. Discovery of program's sources of influence, mechanisms of information exchange, and degree of openness to research reports and the established knowledge base is as much faculties' responsibility as the understanding of the specific effects of course work on teacher performance.

PETOM Experience Summary

The progress of this faculty through program evaluation was collaborative, shared and based on their common understandings about the importance of obtaining data feedback that was useful for program development within the program's conceptual framework. They designed a variety of measures for examining the effect of their own teaching and they struggled to understand the meaning of the data generated by their measures. In collaboration, the faculty designed measures for various program constituencies. In continuing collaborative evaluation work, these measures, constituencies and other contexts will expand in response to faculty's need to obtain useful data for program development. Importantly, they were responsive to evaluative feedback by redesigning their own course work and refining program experiences.

University of California at Santa Cruz (UCSC) In-Service Program Experience

In the past 20 years, teachers have seen an explosion of new ideas and programs for improving classroom instruction. Extensive in-service training initiatives have become the traditional vehicle for conveying new pedagogical strategies. Unfortunately, teachers have typically been viewed as recipients rather than as decision-makers or active participants in staff development programs. Staff development is often seen as "training" or "in-servicing" in which experts teach teachers predetermined instructional methods. This raises an evaluation question, that is, how can veteran teachers be helped to implement new strategies for working with LEP students?

The body of research on in-service training indicates certain characteristics that make it more effective and calls for new ways of looking at retraining teachers. The work of Glickman suggests that teachers need not be trained but rather be given the tools for determining their own instructional priorities. Components that allow teachers to work together and make decisions about planning, implementing, and evaluating instruction (Glickman, 1990) are highly desirable. He found teachers became more thoughtful and resourceful about teaching.

Teacher education and in-servicing strategies such as peer coaching, simulations, demonstrations, performance feedback, interactive journals, and mentoring are currently receiving attention in the literature. These in-service strategies and recent programs like teacher induction enhance the opportunity for professional development within the social, interactive context of teaching activity. Al-

440

though teacher induction programs vary in their design, the overall goals of these programs have been to provide ongoing support and assistance to beginning teachers as they enter the profession, to improve teacher effectiveness, to increase retention in the profession and to promote the professional and personal well-being of new teachers.

Veenman (1984) suggested that novice teachers need both pedagogical assistance and psychological support. This is similar to recommendations from developmental theorists such as Furth (1981) and Vygotsky (1978). They point out that a supportive atmosphere is necessary if learners need to master new and complex thought and action.

As Gherke points out, traditional quantitative methods are inappropriate for trying to understand these relationships. Recent evaluation reports about induction programs suggest that useful data regarding teacher networking, nurturing relationships and complexities of teacher development are available through qualitative methods (Gherke, 1988).

UCSC/Santa Cruz County New Teacher Project

Based on research recommendations on effective in-service training, the University of California/Santa Cruz County New Teacher Project designed an interactive, collaborative program for new teacher support, staff development, and professional growth for veteran teachers. A program titled, the Santa Cruz County New Teacher Project was developed in 1988 by the University of California at Santa Cruz Teacher Education Program in collaboration with the Santa Cruz County Office of Education and seven school districts in the county. In this consortium, communication and collaboration occurred across districts and institutional boundaries. The consortium members determined program philosophy, components, and ongoing evaluation.

During its first three years the project has served 155 first year K-8 teachers with 105 teaching in bilingual classes; 78 of these teachers are graduates of the UCSC Teacher Preparation Program. The Santa Cruz County New Teacher Project supports beginning teachers' efforts to translate what they have learned in their preservice preparation into classroom practice. After nearly three years, the Santa Cruz County New Teacher Project has lost only five of the 155 teachers served.

In this project, the evaluation component was designed by program faculty.and included both quantitative and qualitative methods. Measures, listed in Appendix B, such as standardized interviews, questionnaires, journal entries, videotape feedback, advisor logs, self-assessment forms and weekly observations were systematically used by all project participants. The faculty reviewed data and responded to feedback at weekly meetings of the bilingual advisors and the director. For three years, the program development was guided by evaluation data. What follows is a description of the project and an overview of the findings.

The project serves seven school districts in Santa Cruz County which is a rapidly growing area with an ethnically and linguistically diverse student population of about 33,000 K-12 students. The Pajaro Valley Unified School District, the largest in the county, is one of the most linguistically impacted in California with 40 percent of its students being limited-English proficient. Ninety-four percent of these LEP students are Spanish speaking, and of this group, 77 percent are migrant.

The Santa Cruz County New Teacher Project (SCCNTP), one of 37 pilot projects of the California New Teacher Project, is funded by the State Department of Education and the Commission on Teacher Credentialing. Five exemplary bilingual teachers -- Novice Teacher Advisors -- are the cornerstone of the Santa Cruz County New Teacher Project. The advisors, who specialize in all aspects of bilingual-multicultural education, are hired to work with new teachers for the entire year under the guidance of a UCSC Project Director.

The advisors, whose case load was fourteen bilingual teachers, worked with each new teacher two hours per week both in and out of the classroom setting. The advisors' time in the class was spent doing demonstration lessons in both English and Spanish particularly in language and literacy, observing and coaching, assessing students, videotaping lessons, providing release time, responding to interactive journals, and assisting with problems as they arose. Time outside the classroom was spent on planning, gathering bilingual and culturally relevant resources, problem solving and reflection, and general support and encouragement. By being familiar with the students in the class, the overall curriculum plan, and the class structure and organization, the advisor was able to provide new teachers with context specific assistance.

The overall philosophy of the project is that teaching is complex (Good and Brophy; Shavelson, 1983) and that the process of becoming a teacher involves career-long or life-long learning. The project recognizes that new teachers enter the profession at different developmental stages and with individual needs. In a non-evaluative and supportive manner, the advisors help each new teacher develop an individualized plan to address their specific goals and needs. From week to week the advisor and the new teacher work together to strengthen the new teacher's program with the advisor responding to the new teacher's zone of proximal development. Each new teacher is a member of a team which includes the advisor, the peer resource, and the site principal.

In addition to the ongoing coaching, new teachers received five days of staff development training aimed at meeting the identified needs of the new teacher and their LEP students. The first in-service on classroom management and organization emphasized cooperative learning strategies and heterogeneous groupings. The second inservice on Language Development addressed first- and second-language acquisition, strategies for creating natural language opportunities, methods for integrating language into all areas of the curriculum, and thematic planning. The Reading/Writing Connection focused on ways to create a biliterate learning environment, rich in print and conversation.

The data collected following each session was positive. It indicated that the new teachers felt that they had received information which was immediately applicable and manageable within their first-year context. The weekly coaching that followed supported new teachers in their efforts to acquire new strategies. Two days of release time were spent with the advisor observing exemplary bilingual teachers, discussing the observations, and planning curriculum. A monthly seminar series provided an opportunity for networking and reflection.

Evaluation

As a result of their intensive involvement, the advisors developed a unique and powerful collegial relationship with each of their new teachers. Both midyear and end of the year evaluations from the new teachers used such descriptors as “saint,” “guardian angel,” “friend, and “co-teacher” to describe this relationship. When given an opportunity to describe the most beneficial aspect of their work with the advisor, new teachers,(male and female/ages ranging from 22-50) consistently emphasized emotional support. When new teachers in the Santa Cruz project were asked “Who has been helpful in dealing with the challenges you've faced?” the New Teacher Project was the most frequently named source of support (Drury, 1991). The level of enthusiasm for the project increased throughout the year with teachers repeatedly attesting to the value of the help they received from their advisors and other project personnel.

New teachers and their advisors kept an ongoing interactive journal. The journal entries present the types of questions new teachers pose, the frustrations they face, and the depth of the relationship between the advisor and the advisee. The following excerpt from a third grade bilingual teacher shows her taking charge of her own professional development. “For our next meeting I'd like to meet after school, if possible, to plan a math schedule like you showed us at the in-service. I'd like to bring in some other strands from the math framework like your model does. It is obvious to me that many of the students need additional opportunities for hands on activities."

A second grade bilingual teacher wrote: "I haven't been happy with my writing program. I had great plans at the beginning of the year. I guess where it bogged down was with my frustration with trying to attend to each student's writing needs. It wasn't possible for me to get around individually and actually do much teaching. Because I was disturbed by all these factors I tended to avoid writing, especially revision and editorial stages. Maybe you could demonstrate a writing lesson for me sometime."

After the demonstration lesson he wrote: "I loved the writing lesson. This is exactly what I've been missing, some motivating material to be creative with. They really do need some motivating structure to bounce off from. I guess its asking too much for them to create from a vacuum, which I often have done. As beginning writers, they need to play with vocabulary in this way, one step at a time. I also read your comments on their papers. Makes me see how infrequently I give positive feedback. It's so easy to write "fantastic idea." Why don't I do it more?"

The project director and the advisor had regular contact with the site principals. The principals' perspective about the SCCNTP was collected through both interviews and questionnaires. All of the 30 principals interviewed clearly felt that the most important part of the project was the weekly classroom visits by the advisor. One of the participating elementary principals spoke to the program's effects:

> "The project is supporting new teachers in all the ways that principals would like to but never have the time to do. When I compare the first year teachers of previous years to this year's new teachers, it's apparent to me that this group is much further along in developing their programs. In less than a year, they are doing what it took other teachers three years to accomplish. I attribute this to the close working relationship with the veteran teacher."

These quotes from principals strongly suggest that nurturant relationships within classroom contexts support teacher growth. The advisors' years of experience in bilingual classrooms and their training enabled them to diagnose problems, provide options, and allow the new teachers to remain in control of decision making while being extremely sensitive to the developmental needs of each new teacher.

Again, a principal at a middle school stated: "We saw more growth in our first year teachers than we have ever seen before. By working as a team with a veteran teacher, they are giving the kids in those classrooms a far better education than they would get if that link wasn't there."

One of the unanticipated benefits to veteran teachers at the school site was the spillover effect of the collegiality and collaboration being modeled by the new teacher and the new teacher advisor. When 50 new teachers were asked, "What effect, if any has the New Teacher Project had on your staff?" all but five noted a positive change. The responses included increased networking and sharing of resources, a willingness to share ideas and strategies, a more professional staff, an openness to new ideas; individual staff members have been rejuvenated and motivated to rethink strategies, innovative methods are spreading throughout the school, greater respect for new teachers, and more sensitive to the demands and pressures of first year teachers. Ten of the teachers mentioned that veteran teachers at their school sites often approached them or their advisors to be a part of their sharing or to receive copies of resources the advisors brought. This gave new teachers a boost in self-esteem as they could now be givers rather than always "takers." One new teacher stated in May, "It seemed like the veteran teachers used to run away whenever they saw me coming. But as I've acquired new materials and teaching strategies they often ask about my new ideas. I don't feel like a parasite any longer."

The advisors in the SCCNTP received extensive staff development training in cognitive coaching, clinical supervision, communication skills, the needs and developmental stages of new teachers, and effective strategies for working with LEP students. As part of the evaluation component, the advisors were asked to describe what they'd learned in their two years of working with the project. What follows are examples highlighting the reoccurring themes that emerged.

> I feel a particular benefit from all I have learned about questioning as a means to help teachers think about their teaching and consciously control it through their own decision making. I think that the non-evaluative nature of my relationship with new teachers has allowed me to freely explore the possibilities of novice-veteran teacher interactions. We didn't have to prove anything together. This experience has caused me to develop a better understanding of the dynamics of change as it applies to teachers.
>
> Participating in this Project has me fully convinced that the flourishing of the profession will depend in a large part on the opening of our doors to our peers and that teachers must control the process. It can't be done to them.

By seeing so many different ways to make teaching work I have become more accepting of differences in style and approach. We teach a diverse student population yet I had fallen into the pattern of seeing a few instructional approaches and teaching styles as "best." I have learned that you have to begin with a relationship and then you have the opportunity to dig deeper through questioning to arrive at a closer understanding of the educational decision making that went into creating a learning activity. It is then that you begin a professional dialogue, not with the purpose of persuasion but with an invitation for a thought provoking exchange -- and with this kind of empowerment comes change.

When I faced a problem in helping a new teacher, I could always problem solve with the team of advisors. Together we came up with questioning strategies and new approaches that enabled us and the new teacher to understand better. We were able to "shadow" one another in pairs, peer coaching each other at work with a new teacher. As a member of this collegial network of peers, I was able to grow in my own teaching.

As a member of a team of experienced teachers, working together to support new teachers, I had weekly opportunities to brainstorm, problem solve, share professional expertise and resources with my peers. We felt respected for our talents, skills, strengths, and even for our areas of weakness. Our director always consulted members of the team in decision making and she was able to keep the dialogue open and constructive on a regular basis. This freedom and openness empowered us to take responsibility for our own professionalism. I attribute the success we had in supporting new teachers to this feeling of safety in the community we created among ourselves.

This project is also externally evaluated. The Southwest Regional Laboratory is completing a three year evaluation study of the California New Teacher Project. Preliminary findings conclude that in addition to dramatically affecting retention, high intensity support efforts as provided by the Santa Cruz County New Teacher Project also greatly enhance teacher performance, especially in the area of multicultural education and working effectively with diverse populations. SWRL has also found a positive relationship between the level of support offered to new teachers and the amount of time new teachers actually engage their students in academic learning tasks. In fact, those new teachers who had high intensity support came close to providing as much academically engaged time for their students as do experienced teachers. New teachers in the high intensity models, where they received support and training, reported that the CNTP was important to their success.

Although these quantitative data gathered by SWRL have been valuable, another kind of data was also needed. We needed to study

446

the relationship between the advisor and the advisee in an effort to understand the kinds of interactions that support growth and development in new teachers. Upon reviewing the interactive journals, the advisors were eager to try alternative approaches and/or modify their support based on new teacher needs. The qualitative data in these journals and interviews highlighted the different developmental stages that new teachers go through as well as the type of assistance they sought throughout the year. The advisors had some work to do on their own acceptance and tolerance.

Program responsiveness also extended to the role of the advisors. Most advisors began with the attitude that teachers teach in similar ways with similar methods. In reality, diversity permeated the entire program including evaluation -- different teaching styles, with teachers at different developmental stages, advisors with varying levels of knowledge regarding coaching and biliteracy strategies, principals with different styles and seven districts with different priorities. This openness or embracing of diversity was the heart of the evaluation process. This diversity is illustrated by the questions that continually emerged. How can we help new teachers with limited Spanish who have been assigned to a bilingual class? How can we set up literature groups in a class where management is a major problem? What policy implications does our research design and data have for teacher education and in-service programs? These questions were best answered by collegial teams that gave teachers opportunities to take control of their own professional development and to create non-judgmental strategies for sharing strengths and weakness.

New teachers in the project opened their class up to weekly observation without fear or intimidation. By building the new teachers into the evaluation design, their talents and professional status were acknowledged. This participatory model of evaluation means not "doing evaluation to" program participants but including them in all aspects of the design, implementation, analysis, evaluation and programmatic modification.

Fifth Recommendation

Our fifth recommendation follows from the participatory evaluation used in both the Hawaii and California programs: Engage program participants in program evaluation design. Collaboration on evaluation design produces useful, substantive data that validate program operations and stimulates program responsiveness. A participatory model of evaluation means bringing program participants into the evaluation process from its inception.

UCSC/Santa Cruz County New Teacher Project Summary

The collegial team that was established among all the participants, advisors, new teachers, principals, peer resource, faculty and the director, allowed for and encouraged an open forum for dialogue, continuous reflection, and evaluation. The program focused on advisors being sensitive to the Vygotskian principles in working with new teachers. The multi-modal evaluation with interviews, questionnaires, interactive journals, advisor logs, videotaping, and self-assessment forms were directed toward the constant rethinking and enhancement of the program. Participants viewed the evaluation process as a positive opportunity for professional growth. When the project began in 1988 little was known about the needs of new teachers. It has been through the participation of this collegial team in the evaluation process that we have gained in-depth awareness and insights into the experiences of new teachers serving LEP student populations. New questions arise and the process of evaluation continues.

Summary of Recommendations for Evaluating LEP Preservice and In-Service Programs

In this paper, the teacher education program evaluation literature, the findings of recent research on effective teaching and learning models for linguistic minorities, and the experiences of two programs, one preservice and one in-service, have been discussed. In both programs, a wide range of data were collected in an exploratory manner capturing the perspectives of many of the participants. The programs' evaluation designs sought data about process level issues as well as competency based teacher performance exploring the relationship between what programs actually do with the knowledge base and how teachers develop. Both programs depended on this evaluation data to refine operations and proceeded in a collaborative, activity based manner.

The experiences of these evaluated programs though clearly still evolving reinforce many recommendations in the literature. Building on the literature, the experiences of both programs, and effective teaching and learning strategies for LEP students, we conclude with five recommendations presented as follows and in Appendix A:

448

Recommendations for Evaluation of LEP Preservice and In-Service Teacher Education Programs

1. Identify and document LEP teacher education and in-service programs using program evaluation.

2. Identify and document program responsiveness to evaluation feedback. Evaluation methodology designed to yield relevant, substantive, and useful information for program developers and faculty is most likely to result in program responsiveness to feedback.

3. Use broad based methods and measures to obtain multiple perspectives on LEP teacher education program experience and effect particularly those of the specific communities, cultures, and constituencies served.

4. Design LEP teacher education program evaluation to examine programmatic features important to language development and program responsiveness to new knowledge in the field.

5. Engage program participants in program evaluation design. Collaboration on evaluation design produces useful, substantive data that validates program operations and stimulates program responsiveness. A participatory model of evaluation means bringing program participants into the evaluation process from its inception.

Given the dearth of evaluation activity in LEP teacher training and in-service and the crisis the nation faces in addressing the needs of linguistic minorities, program evaluation demands prioritization. At this time, the linkage between the expanded knowledge about effective education for LEP students and teacher education program development is unclear. Additionally, the relationship between what programs actually do and how teachers develop in them is unreported. The effect of program operations, teacher development processes, and program's linkage to knowledge base are integral sources of information and influence for programs committed to improving teaching practice for LEP students. We need to know much more about these processes and program evaluation as one means to this goal.

Evaluation can be a systematic and regular practice that is an essential component of the program development process, well worth the resources required. By engaging in self-examination, programs themselves enter a developmental process, learning as they teach.

Appendix A

PETOM Evaluation Measures

- Checklists for rating performance competence of the PETOM preservice teachers and controls

- Demographic profiles of ethnicity, age, grade point average, and place of birth

- Preservice teacher interviews about program experience

- Cooperating Teacher interviews

- Rating scales

- Principals' Interviews

- Situational Probes to discriminate between PETOM students and the control group for competence in program emphases such as language development and the use of culturally compatible teaching strategies

- Open Ended Stimulated Recall Interviews about preservice teachers' experiences in the activity of teaching

PETOM Principal's Interview Excerpt

The PETOM graduates "...are very enthusiastic and always willing to leam. They have a good knowledge base and are well equipped with effective teaching strategies although, being new teachers, they are still having difficulties applying them.

They are middle-class teachers dealing with low-income, severely at-risk youngsters, and this is an incredibly hard task. Therefore, they have had to make a lot of adjustments this first year and, in many ways, they are not yet fully equipped. I look at these teachers as slowly evolving, and in two more years they will be top-notch. But, they are never frustrated or depressed. They are lovely people who are trying hard and maintaining very good attitudes despite the many obstacles that working with deprived youngsters brings.

Working on the Leeward Coast is always a very frustrating experience for very new teachers because most of them have not had the hands-on time with these kinds of kids. They have to first experiment with, and weed through a multitude of teaching strategies before finding those that are most effective. These strategies must address all kinds of kids and not just those found at Kamehameha Schools or in town. Most young teachers have not had enough field experience in this area, and it would really help if they spent time out here before graduating from school. ...I think your program is wonderful, but your graduates need to spend more time in areas such as this before they can be truly competent new teachers. These teachers are growing this year and they are always willing to learn. I am sure they will evolve into first-rate teachers."

Appendix B

University of California at Santa Cruz Measures

- Standardized Interviews
- Questionnaires
- Journals
- Videotape Feedback,
- Advisor Logs
- Self-assessment
- Weekly observations were systematically used by all project participants.

University of California at Santa Cruz New Teacher Project

Examples of Program Responsiveness to Evaluation

Based on the data from the first year, the program modified its support to new teachers. The First Year Orientation, focusing heavily on planning, classroom management and bilingual curriculum development was revised to be sensitive to the developmental, needs of new teachers.

In response to data indicating new teachers' increased need for support, the project recruited exemplary, bilingual teachers who could combine their expertise in teaching with nonjudgemental, supportive interpersonal skills.

Based on feedback from new teachers, the seminar series dramatically changed over the three years. Seminars moved from a curriculum emphasis to an open forum, enhancing networking, collaboration, and problem solving.

Advisors began by expecting teachers to teach in similar ways with similar methods. They became responsive to different teaching styles, teachers at different developmental stages, principals with different styles, and seven districts with different priorities.

References

Diaz, R.M. (1983). The impact of bilingualism on cognitive development. In E.W. Gordon (Ed.), Review of research in education (pp. 23-54). Washington, DC: American Educational Research Association.

Drury, R. (1991). Is there life after school? The experience of novice teachers in the Santa Cruz County New Teacher Project. Unpublished master's thesis, University of California at Santa Cruz, Santa Cruz, CA.

Gallimore, R., & Goldenberg, C. (in press). Activity settings of early literacy: Home and school factors in children's emergent literacy. In E. Forman, N. Minick, and C.A. Stone (Eds.), Education and mind: The integration of institutional, social, and developmental processes. Oxford: Oxford University Press.

Galluzzo, G. R. & Craig, J. R. (1990). Evaluation of preservice teacher education programs. In W. Houston (Ed.) Handbook of research on teacher education. New York: Macmillan Publishing Company.

Garcia, E. (1988, Winter). Effective schooling for language minority students. New Focus, (No. 1). (Available from the National Clearinghouse for Bilingual Education)

Garcia, E. (1990). Educating teachers for language minority students. In W. Houston (Ed.) Handbook of research on teacher education (pp. 717-729). New York: Macmillan Publishing Company.

Garcia, E. (1991). The education of linguistically and culturally diverse students: Effective instructional practices (Working paper). Center for Study of Cultural Diversity and Second Language Learning.

Gherke, N. (1988). Toward a definition of mentoring. Theory into Practice, 27(3), 190-194.

Glickman, C.D. (1990). Supervision of instruction: A developmental approach (2nd ed.). Boston: Allyn and Bacon.

Good, T. & Brophy, J. (1987). Looking in classrooms (4th ed.). New York: Harper and Row.

Hakuta, K. (1986). Mirror of language: The debate on bilingualism. New York: Basic Books.

Katz, L., Raths, J., Mohanty, C. Kurachi, A., & Irving, J. (1981). Follow-up studies: Are they worth the trouble. Journal of Teacher Education, 32(2), 18-24.

Little, J. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. American Educational Research Journal, 19(3), 325-340.

McLauglin, B. (1990) Development of bilingualism: Myth and reality. In A. Barona and E. Garcia (Eds.), Children at risk: Poverty, minority status and other issues in educational equality (pp. 65-76). Washington, D.C.: National Association of School Psychologists.

Medley, D.M. (1977). Teacher competence and teacher effectiveness: A review of process-product research. Washington, DC.: American Association of Colleges for Teacher Education. (ERIC Document Reproduction Service No. ED 143 629)

National Council for Accreditation of Teacher Education. (1987). Standards for the accreditation of teacher education. Washington DC: Author.

Native Hawaiian Educational Assessment Project Final Report. (1983). [Executive summary]. U.S. Department of Education: The Kamehameha Schools/Bishop Estate.

Note, M. (1989). Qualitative and quantitative approaches to a preservice teacher education program evaluation. Unpublished master's thesis, University of Hawaii at Manoa, Manoa, HI.

Phillips, M. & Glickman, C. (1991). Peer Coaching: Developmental approach to enhancing teacher thinking. Journal of Staff Development, 12(2), 20-25.

Picard, A.J. & Young, D.B. (1990, September). Methods: teaching and learning principles. The Kamehameha Journal of Education, 1.

Ramirez, J.D., Yuen, S.D., & Ramey, D.R. (1991). Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children. (Contract No. 300-87-0156). Washington, DC: Department of Education.

Sandefur, J.T. (1970). An illustrated model for the evaluation of teacher education graduates. Washington, DC: American Association of Colleges for Teacher Education.

Shavelson, R.J. (1983). Review of research on teachers' pedagogical judgments, plans, and decisions. The Elementary School Journal, 83(4), 392-413.

Showers, B. (1982). Transfer of learning: The contributions of coaching. Eugene, OR: University of Oregon, Center for Educational Policy and Management.

Speidel, G.E. (1990, September). How is PETOM viewed by its graduates? The Kamehameha Journal of Education, 1.

Tharp R., & Gallimore, R. (1988). Rousing minds to life. New York: Cambridge University Press.

Tharp R., & Gallimore, R. (1989). Rousing schools to life. American Educator.

Tikunoff, W. & Ward, B. (1991, April). Competencies of teachers of ethnolinguistically diverse student groups. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Il.

Troyer, M.E. & Pace, C.R. (1944). Evaluation in teacher education. Washington, DC: American Council on Education.

Veenman, S. (1984). Perceived problems of beginning teachers. Review of Educational Research, 54(2), 143-178.

Vygotsky, L. (1978). Mind in society. Cambridge, MA: Harvard University Press.

Wong-Fillmore, Lily (in press). When learning a second language means losing the first. Early Childhood Research Quarterly.

Response to Dalton and Moir's Presentation

Lynn Malarz
Association for Supervision and
Curriculum Development, Virginia

As an educator very interested in professional development and bilingual education, I would like to compliment the authors on their papers and the research that went into writing them. It is an honor to be a discussant and be part of this symposium.

First, let me make a few brief comments about the paper, then I would like to expand on certain aspects of the paper that research has shown to be important components of teacher training. I wholeheartedly agree with the authors that teacher training programs in general are not widely evaluated and it is very unfortunate that the evaluation of LEP teacher training and in-service programs is so neglected at a time when so many states are begging for bilingual teachers. For example, in its August newsletter, the California Administrators Association stated that at the beginning of this year the state would be short 14,000 bilingual teachers.

I can further say that I do not disagree with any of the points that the authors have made in their papers; the papers presented many pertinent issues regarding teacher education. Thus, I invite you to walk down another path with me, one that will expand on parts of the training I see as very crucial components in teacher training. If it is true, as David Berliner (1984) states, that teaching is a constant stream of decisions, and any teacher behavior used is the result of a decision, either conscious or unconscious, then as educators working with teachers, we need to understand how to help teachers make the decision that will promote maximum student learning. By looking at the work of many researchers in the field of education, such as David Berliner Teacher "Executive Processes," Madeline Hunter "Teaching as Decision Making," Robert Goldhammer "Clinical Supervision," Art Costa and Bob Garmston "Cognitive coaching," and many more, (the authors referred to advisers being in-serviced in the cognitive coaching model as well as others), we can begin to understand how an individual can become an autonomous teacher -- teachers who:

- Act With Intentionality
- Generate and Choose from Alternatives
- Use Precise Language
- Take Responsibility
- Monitor, Reflect Upon, and Learn From Experience

- Align Behaviors with Values
- Activate Community

Let's look at Profiling the Teacher (see Appendix A) and the myriad of components that go into the profile. By looking at this, it can be said that "Effective supervision is defined as a set of strategies designed to enhance the teacher's perceptions, decision, and intellectual functions," (Costa, p.17, 1991).

Following are some ways to work with beginning teachers as well as veteran teachers to help enhance their intellectual skills, which contribute to educationally sound decision making.

1. Cognitive Coaching, which I have eluded to, is one way. It is a way that supervisors, advisors, coaches and others can work with teachers to help them with the coaching functions in the four phases of the instructional process (see Appendix A).

2. Peer Coaching is another method that has been shown to increase collegiality and improve teaching. Peer coaching is a confidential process through which teachers share their expertise and provide one another with feedback, support, and assistance for the purpose of refining present skills, learning new skills, and/or solving classroom-related problems.

Any type of coaching system has preconditions that should be considered before implementing the program, such as:

> There must be a general perception on the part of the people involved that they are good but can always get better; they can always improve what they are doing. This general orientation has been found to characterize effective schools.
>
> The teachers and administrators involved must have reasonable level of trust; they must be confident that no one is going to distort the situation in any way.
>
> There must be an interpersonal climate in the school that conveys the sense that people care about each other and are willing to help one another.

Research has also shown that to have meaningful staff development -- programs that become institutionalized -- schools need to have teachers participate in ways that ensure transfer of learning (see Appendix A). Further benefits of coaching programs have also been documented:

- Better understanding of teaching
- Improved self-analysis skills

- Improved sense of professional skill
- Renewal and recognition
- Increased sense of efficacy
- Increased collaboration/collegiality
- Improved teaching performance
- Increased student growth/development

Now one can ask, "what is the purpose to all this?" I know it has been verbalized these last few days at this symposium, but let me again state what I feel is necessary in bilingual education -- teachers need to increase their responses to students that will support and extend student thinking and learning. By helping teachers learn different strategies (cognitive coaching, peer coaching, etcetera.), higher order thinking skills can be attained.

Research has found the manner in which teachers respond to students has great influence on the student. Different researchers have documented (Lowrey, 1990) that the way teachers respond has a greater influence on students' thinking than what the teacher asks or tells them to do. Students are constantly anticipating how their teachers will respond to their actions. Thus, the way teachers respond to students seems to exert greater influence than the teachers' questioning. It has also been found that teachers' responses have a great deal of influence on the development of students' self-concept, their attitude toward learning, their achievement, and their classroom rapport.

Let me quickly review response behaviors of teachers -- teacher initiated questions and directions that elicit thinking and learning (See Appendix A).

Last, I would agree with the authors that teacher training and in-service needs systematic and regular evaluation. My hope would be that more programs not only have approaches and processes similar to UCSC/Santa Cruz County New Teacher Project and the University of Hawaii's Preservice Education for Teachers of Minorities but also incorporate the research that is constantly emerging on higher order thinking skills. Skills that can enhance a student's learning; skills that will help the student become a life-long learner. The Greeks had a word for it: Paideia:

> A society in which learning, fulfillment, and becoming human are the primary goals and all its institutions are directed toward that end. The Athenians designed their society to bring all its members to the fullest development of their highest powers. They were educated by their culture -- by Paideia. Self-development and the promotion of lifelong learning is the "central project" of society.

Appendix A

THE MANNER IN WHICH TEACHERS RESPOND TO STUDENTS HAS GREAT INFLUENCE ON THE STUDENT.

TEACHERS NEED TO ENHANCE COGNITIVE LEVELS OF CLASSROOM INTERACTION.

Gathering and Recalling Information (Input)

To cause the student to input data, questions and statements are designed to draw from the student the concepts, information, feelings or experiences acquired in the past and stored in long or short term memory.

Some verbs that may serve as the predicate of a behavioral objective statement are:

completing
counting
defining
describing
identifying
listing
matching
naming
observing
reciting
selecting
scanning

Questions:

"How does the picture make you feel?"	Describing
"The Mexican houses were made of mud bricks called what?"	Completing

Making Sense Out of Information Gathering (Processing)

To cause the student to PROCESS the data gathered through the senses and retrieve from long- or short-term memory.

Examples:

- synthesizing
- analyzing
- classifying
- comparing
- contrasting
- distinguishing
- experimenting
- categorizing
- explaining
- grouping
- inferring
- making analogies
- organizing
- sequencing

Questions:

"Compare the strength of steel to the strength of copper."	Comparing
"How can you arrange the rocks in the order of their size?"	Sequencing
"How are pine needles different from redwood needles?"	Contrasting

Raising Questions to Higher Levels

	Input	**Processing**	**Output**
1.	How many of you are buying milk today?	Why are you buying milk today? Are there more students in Mr. Jones' room than in ours who are buying milk? Why are there fewer milk buyers in our room?	What do you think would happen if nobody bought milk anymore? What do you think would happen if all the children in the world had all the milk they needed? Could you give some examples of countries where this is the case?

2.	What was the weather like yesterday?	How does the weather today compare to the weather yesterday?	What do you think the weather will be like tomorrow?
		Why is our weather so different today?	What can you say about cities which have weather like ours?
		How does the weather in Washington, DC compare with the weather in Tokyo?	

A Model of Intellectual Functioning

Input	**Processing**	**Output**
Intake of data through the senses	Making sense out of the data	Applying and Evaluating
Recalling from both short- and long-term memory		Metacognition

Response to Dalton and Noir's Presentation

Victoria Jew
California State University, Sacramento

The focus of my discussion will be on Recommendation Four: that we design teacher preparation program evaluation to examine programmatic features that are important to the development of the candidates and that we examine a program's responsiveness to new knowledge. I think that we need to focus on two areas of this recommendation because of the nature of this particular field: one is training resiliency because of the school sites that our bilingual education candidate will eventually work in, the other is the responsiveness to the changing demographics of the LEP students -- this is in addition to the general responsiveness to the emergence of new knowledge.

Let me focus on the first point: the condition under which our candidates ultimately have to teach. Bilingual education, as we all know, is still a controversial field. It is controversial because of the specific characteristics of the knowledge base, the population it serves, and the broader social and political climate under which the education of limited English proficient students operates. This condition still exists in a lot of the school settings in which our students will eventually work.

Our graduates are often placed in non-nurturing or even hostile school environments because bilingual education and other approaches to teaching limited English proficient students force schools to address language and culture issues and because they elevate the needs of non-mainstream students into an educational priority at these school sites.

To add to this condition is the controversial nature of our knowledge base. Although we have a substantial knowledge base about language development and effective practices for LEP students, a substantial body of this knowledge is controversial because it may be contrary to the belief system of the school culture of the sites in which our graduates will have to work. Whereas in non-controversial areas of this knowledge base, training programs can merely address the program design in terms of effective implementation, such as: assisting the candidates in developing a professional perspective or working on their competency to bridge theory and practice. But in controversial areas, we really need to address additional program design features to look at the effect of training when it comes into contact with a hostile or non-nurturing environment.

Let me give you some examples about the nature of the body of knowledge that I am referring to. If you are looking at sheltered English, whether you are looking at its theoretical and philosophical perspective or its approach and strategies, it is not controversial. So in training, it is something that can probably be dealt with by employing regular measures for increasing the effectiveness of training. I suspect language choice for instruction in a multilingual classroom will be just as non-controversial. But if you are to look at approaches such as substantive and substantial use of the primary language for instruction, criteria for transition, or language maintenance, then you will run into areas where the specific school culture might be completely not in congruence with the knowledge base in which the new bilingual teacher has been trained. Another area of the knowledge base that was mentioned in the paper: I can more or less predict that the work Lily Wong Fillmore is doing right now that warns of the harmful effects of early English immersion for young children will be controversial when it gets to a school setting.

Let me share with you one of my greatest frustration as a teacher trainer which is also the frustration many of us in teacher training share. We prepare bilingual teachers who appear to be well-trained as they complete the training program. They appear to have at least a well defined philosophical perspective; they have developed a sound professional perspective; they show beginning level of competency in methodology, strategy, and practice that are effective in classrooms for LEP students. But within two years after graduation, when we visit them in their teaching setting, a good number of these former students whom we have trained look no different from others at the site who have not been trained. They have become socialized into this particular environment, and we see little sign of the training they have received.

For some of these students, the rhetoric of bilingual education remains, but there is no reflection of that particular perspective in their classroom practices. Others completely abandon what they have gained in the training program and look just like everyone else. Still others even work against their training, using the knowledge base they were trained in to attack bilingual education. I am wondering, for those who have abandoned the skills that they developed when they were faced with a hostile environment, whether the design of a training program can focus on preparing the candidates to face the special kind of challenge that bilingual educators often have to face: the lack of support, the lack of material, the lack of commitment, the lack of resources and the lack of status in schools.

Some of the enhancement of the candidates can be addressed when you look at how we bridge theory and practice, how we assist students in practice teaching, how we coach students. But there are other areas that need something more than these types of improvement in training. Therefore, I really think that data needs to be collected to explore ways in which the process of teacher development can incorporate the kind of training and support that would enhance the resiliency for the special kind of knowledge base that we work from; the belief system and the commitment to an effective education for language minority students with which a training program prepares, trains, and empowers bilingual educators.

The second area I want to give focus to is the responsiveness to new knowledge. Within the consideration for new knowledge, we need to consider responsiveness to the changing demographics of the LEP student population in schools. We need to evaluate how effectively training programs continue to modify programs as they collect data about emerging populations of LEP students. Let me give you some examples of the kind of frustration that we have in the field of bilingual teacher training, particularly in the training of teachers for working with Asian-Pacific language populations. There is an absolute dearth of data or research on this particular population in terms of bilingual education or second language education in the U.S. This is also true for any language other than Spanish. In many ways, a lot of us, who are working in training programs that train for other languages, experience a tremendous amount of difficulties in presenting data that can be considered robust.

In addition, there is a certain degree of what may be considered capriciousness in the way we determine what is generalizable or not generalizable to other LEP populations from the data we have about Spanish speaking students. Some of us are quite surprised by the type of interpretation that others in bilingual education render when it comes to non Spanish languages. One example, for instances, is the question of literacy transferability. In California, there has always been a focus on the extent of transferability of literacy between L1 and L2. This is all well and fine for two languages such as Spanish and English, not very distant in many ways. Then when you have candidates in training who are from other language backgrounds, including those from languages with extremely different orthographies such as Chinese, the capricious manner in which one trainer determines that there is zero transferability and another determines to claim extensive transferability without any attention to the obvious differences can be most confusing.

Another recent surprise is the Summary of the Longitudinal Study by Ramirez in which a statement was made that the results of the study cannot be generalized to other language groups, because research indicated that other language background students acquire English differently. There is no body of research that we are aware of about how other language groups acquire English. Given the lack of data for interpretation, this statement in the summary again is open to capricious interpretation that can be both surprising to people who are speakers of these languages and confusing to candidates in training.

Finally, we really need to start to take a look at how we can continue to incorporate a pluralistic perspective in the way we train bilingual teachers. A lot of the paradigms for training or in classroom practices in bilingual education or in language development education have been set based on the experiences, practices, and data of the Spanish-speaking population. Thus, our programs need to give focus to evaluation that examines how well we continue to include and how well we continue to change in order to address changing demographics.

INDEX

Volume I

RESEARCH PAPERS

Baron, Joan Boykoff 187
Dalton, Stephanie 415
Damico, Jack S. 137
French, Russell L. 249
Garcia, Eugene 383
Ginsburg, Alan L. 31
Moir, Ellen 415
Oller, John W., Jr. 43
Ortiz, Alba A. 315
Popkewitz, Thomas S. 287
Walters, Joseph 1

DISCUSSANTS

Davidson, Fred 125
Clements, Barbara 377
Figueroa, Richard A. 243
Habermann, Mary Jean 235
Jew, Victoria 453
John-Steiner, Vera 19
Kawakami, Alice J. 273
Koretz, Daniel 281
Malarz, Lynn 447
Met, Myriam 131
Migdail, Sherry R. 353
Navarrete, Cecilia J. 183
O'Malley, J. Michael 173
Teele, Sue 23
Willig, Ann C. 343

PANELISTS

Allegro, Annalisa 363
Izquierdo, Elena 373
Romero, Migdalia 367

Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues:

FOCUS ON EVALUATION AND MEASUREMENT

VOLUME 2

United States Department of Education
Office of Bilingual Education and Minority Languages Affairs

BEST COPY AVAILABLE

Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues:

FOCUS ON EVALUATION AND MEASUREMENT

Washington, D.C.
September 1991

VOLUME 2

United States Department of Education
Office of Bilingual Education and Minority Languages Affairs

Published August 1992

"The papers in this publication were commissioned by the Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education. However, these contents do not necessarily represent the policy of the U.S. Department of Education, and endorsement by the Federal Government should not be assumed."

FOREWORD

With this publication OBEMLA adds twenty research papers to the ten presented at the First National Research Symposium in 1990. The focus of these papers, delivered at the Second Symposium on LEP Student Issues, is especially timely. Evaluation -- understood not only as a technique but, more important, as a habit of thought -- is still in its infancy. This is as true of education as it is of business or social services. One has merely to read the daily newspaper regularly to become aware that evaluation is a recurring preoccupation in any institution -- whether a Fortune 500 corporation or a private academy or a drug rehabilitation center -- that convenes people around a shared task. Evaluation enables us to discover certain facts about the past and the present -- what works and what does not. But that is not enough. Evaluation must also reveal to us the how's and why's so that we can make judgments about the future, so that we can deliberately choose our next steps.

At last year's symposium I noted the importance of research, from which I expect both the theoretical framework and the factual grounding of effective second language learning processes. To this affirmation, I want to add another: ultimately, the conclusions of research must be accessible to the people who make policy, who teach, who design curricula, and, yes, even to the people who seem the furthest removed from academia -- the plain ordinary parents of plain ordinary language minority students. My words are not intended to bash "pure" scholarship or "ivory towers"; above all, they do not dismiss those who study and think and analyze and construct new theory. On the contrary, I respect the work of scholars and value their contribution to a task that is large enough to utilize the diverse talents of all of us. But I do mean to underline a central fact: if the knowledge and the understanding created by research do not ultimately enlighten the publics I mentioned, the field will never reach the breakthrough insights and decisions demanded by the mammoth needs of students. In relation to the topic at hand, evaluation, the broad accessibility of research findings is key to the educational regeneration we seek. I challenge the research community, therefore, to be inventive about the interpretation and transmission of findings.

I know that the research reported in these papers will make significant contributions to the thousands who work with and for language minority students. We at OBEMLA will surely take them to heart. I am proud of OBEMLA's role in promoting research and grateful to Dr. Carmen Simich-Dudgeon and her staff for their efforts in planning and conducting the symposia.

Rita Esquivel
Director
Office of Bilingual Education
and Minority Languages Affairs
U.S. Department of Education

(Note: On May 30, 1992, Nguyen Ngoc-Bich assumed the role of OBEMLA's Acting Director. Rita Esquivel resigned her position as OBEMLA's Director to resume her career with the Santa Monica-Malibu Unified School District, California.)

ACKNOWLEDGEMENTS

The OBEMLA Research and Evaluation Staff (RES) wishes to express its appreciation to the authors of the commissioned papers that appear in this volume. These papers were presented at the Second Research Symposium on Limited English Proficient (LEP) Issues held in Washington DC, September 4-6, 1991. The authors are: Eva Baker, University of California, Los Angeles; Kurt Geisinger, Fordham University, Bronx; JoAnn Canales, North Texas State University; Thomas M. Haladyna, Arizona State University, Tempe; Walter Secada, University of Wisconsin; Beth Warren, and Ann S. Rosebery, Technical Education Research Center, Massachusetts; Liz Hamp-Lyons, University of Colorado; Peter Negroni, Springfield Public Schools, Massachusetts; John Steffens, University of Oklahoma, Norman; and Carl Grant, University of Wisconsin, Madison.

The collection of papers in this volume represents the United States Department of Education's continued commitment to provide a forum for the exchange of information and discussion of diverse points of view regarding the education of LEP students.

RES also wishes to recognize and thank the following symposium discussants and panelists: Lorraine Valdez Pierce, Center for Applied Linguistics; Peter Byron, New York State Education Department; Michele Hewlett-Gomez, Sam Houston State University; Larry Rudner, ERIC Clearinghouse of Tests, Measurement and Evaluation; Julia Lara, Council of Chief State School Officers; Robert Rueda, University of Southern California; Gary Hargett, University of New Mexico; María Pennock-Román, Education Testing Service; Raj Balu, Chicago Public Schools; Jesus Salazar, Los Angeles Unified School District; Tomi Berney, New York City Public Schools; Robert Martinez, University of New Mexico; Penelope L. Peterson, University of Michigan; Mary M. Lindquist, National Council of Teachers of Mathematics; Ron Rohac, San Bernardino City Unified School District, California; Sau-Lim Tsang, ARC Associates, Inc.; Denise McKeon, George Washington University; Joy Peyton, Center for Applied Linguistics; Virginia Collier, George Mason University; Rosita Galang, University of San Francisco; Margarita Calderon, University of Texas, El Paso; and Li-Rong Lilly Cheng, San Diego State University.

RES is especially appreciative of the collaboration provided by the U.S. Department of Education's Office of Educational Research and Improvement (OERI), and particularly by Dr. René Gonzalez, OERI's Center Monitor for the National Center for Research on Cultural Diversity and Language Learning at the University of California, Santa Cruz.

Finally, my special thanks to Alex Stein, Cindy Towsner, and Cecile Kreins of my staff, for assisting authors, discussants, and panelists during the paper submission process. My sincere appreciation to Timothy D'Emilio and Sharon Coleman of my staff for their hard work in coordinating the publication of Volume I and Volume II, the proceedings of the Second Research Symposium with a focus on evaluation and measurement.

Carmen Simich-Dudgeon
Director, Research and Evaluation
Office of Bilingual Education and Minority Languages Affairs
US Department of Education

CONTENTS

Volume II

FOREWORD iii

ACKNOWLEDGEMENTS v

INTRODUCTION xi

Issues in Policy, Assessment, and Equity 1

Eva L. Baker

Discussants:

Lorraine Valdez Pierce 19
Peter M. Byron 27

Testing LEP Students for Minimum, Competency and High School Graduation 33

Kurt F. Geisinger

Discussants:

Michele R. Hewlett-Gomez 69
Lawrence M. Rudner 81

Innovative Practices in the Identification of LEP Students 89

JoAnn Canales

Discussants:

Julia Lara 123
Robert Rueda 127

Test Score Pollution: Implications for LEP Students 135

Thomas Haladyna

Discussants:

Gary Hargett 165
María Pennock-Román 173

LEA Title VII Program Evaluations

Panelists:

Raj Balu 183
Jesus Salazar 187
Tomi Deutshc Berney 193

Discussant:

Robert Martinez 199

Evaluating Mathematics Education of LEP Students in a Time of Educational Change 209

Walter Secada

Discussants:

Penelope L. Peterson 257
Mary Lindquist 267

Science Education as a Sense-Making Practice: Implications for Assessment 273

Beth Warren and Ann S. Rosebery

Discussants:

Ron Rohac 305
Sau-Lim Tsang 311

Holistic Writing Assessment for LEP Students 317

Liz Hamp-Lyons

Discussants:

Denise McKeon 359
Joy Kreeft Peyton 365

A Superintendent's Evaluation of Teacher Education Reforms 371

Peter J. Negroni

Designing an IHE Teacher Training Program for Specific LEP Student Instructional Needs 393

John E. Steffens

Discussants:

Virginia Collier 417
Rosita G. Galang 423

Educational Research and Teacher Training for Successfully Teaching LEP Students 431

Carl A. Grant

Discussants:

Margarita Calderon 457
Li-Rong Lilly Cheng 473

INDEX 477

INTRODUCTION

This is Volume II of two volumes that contain the proceedings of the Second National Research Symposium on Limited English Proficient Student Issues. The Symposium represented a collaborative effort between the Office of Bilingual Education and Minority Languages Affairs (OBEMLA) and the Office of Educational Research and Improvement (OERI) and was held in Washington DC, September 4 through 6, 1991.

The general theme of the papers in these volumes is evaluation and measurement. It is an effort on the part of OBEMLA to promote the dissemination of state-of-the-art information regarding key issues in the education of school-age students of limited English proficiency (LEP). Specifically, the papers discuss both the theory and its application in the area of educational evaluation and measurement and the role of assessment in terms of accountability and program improvement at the federal, state, and local levels. In addition, evaluation and measurement issues in other areas are discussed. For example, the evaluation of teacher education programs, both at the preservice and in-service levels and the evaluation of curricula, e.g., science and math, in view of advances in these and other fields are topics covered in these volumes. Other topics, including the applications of foreign language testing to second language learning are discussed,as is research on multiple intelligence, its present and future impact on changes in the way we envision the field of evaluation and measurement,and its initial applications to LEP student and program evaluation.

We believe that dissemination of innovations in evaluation and measurement are at the core of the school reform movement. The papers in this volume we hope will act as catalysts to dialogue between practitioners and researchers about alternative assessment theories, methods, and strategies, and their potential application to the assessment of LEP students' language and subject matter knowledge. Furthermore, we hope that discussion will expand to include issues of program evaluation and improvement. Alternative assessment practices, including portfolio assessment and holistic writing assessment, are innovative trends in evaluation and measurement whose time has come. We encourage further study of these innovations and discussion of diverse points of view on the merits and constraints of these methods in the education of LEP students.

The remainder of this section consists of brief summaries of the main issues discussed in each paper.

Eva Baker provides an overview of the literature on innovative assessment research and its applications as they relate to national education policy. In "Alternative Assessment and National Education Policy," Dr. Baker argues that the role of assessment is key in educational reform and in the accomplishment of the President's National Education Goals. She suggests that prevailing view is that drastic changes in the nature of assessment, from molecular, multiple choice formats toward more complex, meaningful, and integrative performance tasks, will result in improvements across the full range of educational activities. Alternative assessment, she states, challenges our current views of the curriculum, of teaching practices, and of the presentation of student achievement information to policy makers and to the public.

"Testing LEP Students for Promotion, Minimum Competency and Graduation," by Kurt Geisinger, reviews the current status and content of typical minimum competency examinations as used for making high school graduation decisions across many states and school districts of our country. The author contends that these examinations have generally called for the students to demonstrate basic subject matter mechanics and/or the application of what has been called survival skills for adult living. Dr. Geisinger suggests that the states have no consistent manner in which LEP students are assessed on statewide or district-level minimum competency examinations and discusses some of the methods by which states make these decisions. Issues bearing on the evaluation of minimum competency examinations are discussed, especially reliability and validity. With regard to validity, instructional and curricular validity are described as especially important as they were seen as critical to the *Debra P.* judicial case in Florida. The concepts of differential curricular validity and differential instructional validity are introduced from the perspective of LEP students including suggestions for setting standards for minimum competency.

Jo Ann Canales' paper, "Innovative Practices in the Identification of LEP Students," was aimed at providing an information base regarding current identification practices used by the Texas state department of education as well as measures recommended for assessing integrative use of language skills. The author suggests a way to systematically identify Limited English Proficient students using multiple, alternative criteria and offers a paradigm that can assist state departments of education to collect consistent data regarding the students in need of English language assistance.

The model proposed in this paper is called the English Language Assistance Profile Chart because the process involved exceeds the traditional practices of identification and can be used to make decisions for placement and exit, as well. Use of this model, states the author, consolidates the gathering of information for practitioners and enables them to make informed decisions regarding the needs of linguistically different children.

The concept of test score pollution is introduced by Thomas Haladyna in his paper, "Standardized Tests Have a Multitude of Interpretations and Uses." Test score pollution, he states, is a condition that affects the validity of these interpretations and uses. The author presents the problem of test score pollution in the context of achievement testing, speculates about its origins, provides evidence of its complexity and severity, and addresses the implications of test score pollution for limited English proficient students.

Walter Secada's paper, "Evaluating Mathematics Programs for LEP Students," explores issues related to the curriculum, the quality of academic instruction that LEP students receive, the kinds of assessments that support quality instruction, and the implicit theories of student cognition, of learning, or of curriculum that were embedded in these practices.

Dr. Secada states that the academic subject areas in general, school mathematics in particular, are undergoing a period of intense scrutiny and reform, comparable in scope to the post-Sputnik crisis which gave rise to the New Mathematics of the 1960s. He suggests that the mathematics education reform movement has failed to pay serious attention to the education of diverse learners. The proposed mathematics standards, suggests the author, do not include checks to ensure that they will, in fact, apply to everyone nor that the practices that are promulgated meet the diverse needs of LEP students.

In "Science Education as a Sense-Making Practice: Implications for Assessment," Beth Warren introduces a scientific, sense-making approach to science education for language minority, LEP students and explores the implications of this approach for assessment. The view put forward, which was referred to as scientific sense-making, is grounded in both cognitive science and sociocultural views of learning. Drawing on concrete examples of scientific sense-making in language minority classrooms, Dr. Warren then explores possible contexts of assessment that arise in a sense-making culture and the role assessment can play. The contexts of assessment include students' talk, texts, and scientific activity. The role of assessment in a sense-making culture extends beyond monitoring to promoting student learning. In conclusion, the author suggests implications of a sense-making view for improving science education and assessment, paying particular attention to issues of teacher development and empowerment.

In "Holistic Writing Assessment for LEP Students," Liz Hamp-Lyons focuses on writing assessment. Dr. Hamp-Lyons discusses writing assessment in terms of both accuracy and appropriacy.

The author describes the available types of holistic writing assessment, placing these on a continuum toward increasing humanism. She goes on to consider a wide range of contexts, from individual classroom evaluation to program-wide and statewide performance and accountability testing and proposes, describes and illustrates one specific form of holistic writing assessment -- multiple trait assessment -- which offers the most humanistic and meaningful assessment of the writing of LEP learners currently available. Dr. Hamp-Lyons continues by describing how psychometric reliability and validity can be assured for multiple trait measures and closes by briefly referring to portfolio assessment, and describing the steps presently being taken to improve accuracy and accountability of portfolio evaluations.

Peter Negroni examines the reform efforts in teacher education in the United States within the context of the broader restructuring and reform efforts that are taking place in this country. His paper, "A Superintendent's Evaluation of Teacher Education Reforms," chronicles teacher education reform efforts since "The Nation at Risk" and describes three major approaches that have been used: state driven reform, teacher education reform, and school centered reform.

The author states that major transformations must take place in public education along four dimensions: organizational, pedagogical, social and attitudinal, and political -- in order for our schools to effectively prepare youngsters for this new world and suggests that teacher education reform, while critical as part of a total systemic change process, will not be enough to produce the kind of school we will need for the twenty-first century.

In "Will the LEP Train Reach its Destination: Designing an IHE Teacher Training Program for Specific LEP Student Instructional Needs," John E. Steffens discusses new paradigms in teacher education for teachers of LEP students and provides the background and preliminary knowledge base to substantiate a call for action. Dr. Steffens argues for the need to paying attention to LEP students needs in educational reform and restructuring, particularly in order

479

for these students to met the National Education Goals. The author describes steps that need to be taken now to accomplish the suggestions outlined in the paper.

Carl Grant's paper, "Successful Innovations in Teacher Education Programs," examines the research on teacher education particularly as it relates to preservice and in-service teacher preparation of teachers who work with limited English proficient students. Dr. Grant highlights successful program patterns and innovations based on research for preparing teachers to work with LEP students. A discussion of the criteria used to determine programmatic success is suggested.

Carmen Simich-Dudgeon
Director, Research and Evaluation
Office of Bilingual Education and Minority Languages Affairs
US Department of Education

Issues in Policy, Assessment, and Equity

Eva L. Baker
University of California, Los Angeles

National educational reform presents an unprecedented opportunity to combine our boldest policy options, the best technical knowledge, and American concerns about equity and fairness. To date, the intent of the National Education Goals supports the policy goal of high quality education and restoration of American competitiveness. The six Goals (see Appendix) also refer to challenge and accomplishment as required of "all" students. Particularly, in Goal One, focused on children's readiness for school, the explicit acknowledgment of the importance of health care, early education, and parental guidance push the boundaries of educational reform far beyond the schoolhouse door.

In the last year, efforts have been mounted on the national scene to convert the National Education Goals into policy. The appointment of the National Council on Education Standards and Testing, specifically commissioned to focus on Goal Three, has deliberated on the following questions:

> Is it desirable and feasible to have National Standards of Education?
>
> Is it desirable and feasible to have a National System of Examinations?
>
> Can these policies be implemented while respecting the traditions and legal constraints of local educational control and authority?
>
> What structure or mechanisms should oversee the development of these Standards and Assessments?

Within a six month period, a panel of 32 individuals -- senators, governors, congressmen, administration representatives, educators, and other public figures considered these questions.

The assumptions underlying a set of national education standards, in part, grow from the observation of the successes of other educational systems, particularly those of our trading partners in the Far East and in Europe. Most of these countries have some form of national curriculum since education is a centralized function. Many of these countries have histories of national examination systems, where individual students received certificates of proficiency or passports to higher education linked to their specific educational accomplishments. Despite evidence and argument that the infrastructures of these countries support education in a far different manner than in the United States, prestige of the teaching profession, for instance, and that certain cultures support a set of explicit and early decisions about the track a child will take in school and in life, there are a great many other concerns about importing educational models into the particular United States context.

The United States differs in important ways from most of the countries we believe to have exemplary educational systems. First, the U.S. is much more diverse -- in economics, in culture, in first languages spoken -- than any of our competitors. Second, the population is greater -- more children are in school at a grade level or two than the total population of countries we are supposed to emulate.

Third, structural nature of poverty among some groups works against a school-based educational reform strategy used in other nations.

And last, the United States possesses, as almost all social critics, foreign and domestic, have noted, a set of values in tension that define many important attributes of American life. Whether from a historical or literary approach, these values seem to suffuse American life and at once provide the context for much of the conflict played out in successive policy options.

Values

Think of America, or read social commentary from 150 years ago to the present and become confronted with the idea of fairness. Fairness is a proposition subscribed to by all but defined differently. In an educational context, it is interpreted in terms of opportunity as well as in terms of outcomes. Schools must provide equal opportunity for learning; legal precedent has held that certain tests are biased (unfair) if particular racial and ethnic groups fail in disproportionate numbers. Fairness eludes the schools and, as our student bodies become more diverse, the schools must find ways to deal with children from cultures, languages, and expectations that mainstream America barely understands, if at all. Fairness is also a matter of financing and, to this point, a national educational plan must address inequities in educational resources.

Pluralism is another key value in America, borne of our immigrant history. Freedom of expression, tolerance for modes of living that differ, and respect for individuals from all backgrounds are cultural mantras. In the educational arena, the boundaries of pluralism are being pushed by arguments for multicultural curricula -- even separate curricula for individual groups. Changing the game from how we interact with one another to a differentiated content in the curriculum is certain to present perplexing policy options in the future.

When Americans describe themselves, it is often in terms of individualism, a value related to pluralism but with a very different slant. We value a person's right to define his or her own personal goals and to pursue them. We like idiosyncrasy and celebrate individual achievements. Our educational system reflects this value by revering an individual teacher's right to conduct classroom activities by his or her own lights. We give students many choices -- of topic and of courses -- so they can ideally fashion part of their own educational experience.

We also believe in the idea of self-renewal, that people can start over, in 80s language, "reinvent" themselves. Thus, any course of action can be changed, failures, ideally can be overcome; class membership is not a permanent state.

Competition, the need to win or be best, is at the heart of the American psyche. It is exemplified in our economic system, in our obsession with sports and awards, in school with the emphasis on grades and comparisons, state with state, or child with child. It is, in part, an explanation for how the psychometrics of the twentieth century developed to differentiate performance among people rather than to describe its characteristics.

Finally, Americans also believe in community, in the importance of our neighbors, in helping, and in providing aid and mutual support.

It takes little pondering to discern that these values create clusters of tension as emphasis upon them differs by group, by goals held, and over time. Yet, it is precisely these values, and adherence to them by different participants in educational policy, which frame the conflict about national educational standards and testing.

Standards

The term standards in education has meant to define the level of desired performance. In the current debate, educational standards have come to describe, in part, what content and skills a student was supposed to know. This change in usage is, in part, a public relations move, to convey the notion of "high standards." But it is also relevant to the issue of local control. Instead of discussing national curriculum, a topic sure to draw heat from a variety of sources, the term standards is somewhat sanitized by its ambiguity. However, when content standards are discussed, that is, what is expected of a student in mathematics or science, it is in fact curriculum goals that are really being discussed.

There is also considerable discussion about the strategy by which standards get enunciated and ratified. All agree that the standards should be consensual. They should be developed by representative groups of scholars and practitioners and reviewed by teachers, embraced by policy makers, and so on. Some believe that it is best to begin the process from the end, by creating examples of performances that students should exhibit and derive the standards from this set of performances. In an earlier era, this strategy was called backward chaining and worked when one had a good idea of what the goal was to be in the first place. In subject matters where consensus may be difficult to find, for instance, in literature or social studies, this strategy seems less sensible.

The model used by policy makers in the recent discussion of national standards has been the standards developed by the National Council of Teachers of Mathematics. These standards define, in fairly global terms, what is expected of students in mathematics. The standards are notable because they emphasize problem solving and applications of mathematical thinking, such as estimation and measurement. Because these standards were developed with contributions and participation from many major players in mathematics education, they are often held up as an example of what should occur in the other subject matter fields identified in Goal Three of the National Education Goals: language arts, geography, history, and science. Underway at the present time are consensus processes in history and science. Additional efforts, focused on developing common objectives for the National Assessment of Educational Progress (NAEP), are in various stages in language arts and geography, as well as art.

In the development of the report Raising Standards for American Education, the National Council on Education Standards and Testing identified not only content standards, described above, but also performance standards -- designed to provide a common language for describing proficiency. Most controversial and the topic of some acrimonious debate was the topic of delivery standards. Simply stated, delivery standards were to describe the desirable characteristics of schools and educational systems. The purposes of such description were to assure that schools provided reasonable opportunities for students and to permit analyses and explanation of student outcomes, appropriately conditioned by their educational experiences.

National Systems of Examinations

The standards issue pales in comparison to the issue of a national examination system. Proponents of such a system argue for it from a variety of platforms. Some see its value in operationalizing standards for accountability purposes. They see the function of examinations in terms of sanctions for poor performance and rewards for achievement. Others believe, again using the accountability line of argument, that common examinations will permit comparisons among children, schools, and states, and drive, via the value of competition, performance upward. For these proponents, the form of the examination makes little difference, although most agree it should reflect curriculum and standards.

For others, the power of a national system of examinations inheres, in part, in changing dramatically the form of tests administered to students. At issue is the effect of multiple choice tests on the quality of education. Although everyone would be quick to acknowledge that any test has a reductionist function, multiple choice tests have come in for a strong share of criticism. They are blamed for the piecemeal way teaching and learning occurs (presumably modelling from the format of the test) and for hours spent away from real instruction and focused on test taking skills.

The alternative proposed is a seemingly new form of assessment -- assessment that depends upon students completing longer term tasks, such as essays or projects, and engaging in multiple steps. Instead of multiple choice responses, the students construct their answers and display their proficiencies either in their own performance, such as giving a speech, or in a product they have made, such as an essay or a videotape. One characteristic of these alternative assessments is that they are supposed to be intrinsically motivating, a kind of 1990s relevance. They also may encourage the integration of knowledge across the disciplines.

Thus, alternative assessments focus on students' performance on tasks that require extended time, complex thinking, and integration of subject matter learning (Baker & Linn, 1990; Shavelson, 1990; Torney-Purta, 1990). For leaders in the research and policy communities, the recognition that measures of educational achievement should reflect the complexity of learning has created enormous opportunity to reform education through providing a focus on curriculum, staff development, and instructional improvement (Ambach, 1991; California Assessment Program, 1991; Baron, 1990; Resnick, 1990).

Examples of alternative assessments might be as common as an essay examination or might include tasks such as the following:

1. Situate an aquarium in the school cafeteria.
2. Make a pinwheel (sailboat, or kite) and explain how it works.
3. Create a work-readiness portfolio with evidence of writing, teamwork, technology use.
4. Design, justify, and estimate costs for recreational facilities for your neighborhood.

It is clear that to judge the quality of such tasks, observers or raters must be trained to use specific scoring rules and to demonstrate their ability to do so with reliability, validity, and without bias.

Alternative assessment is promulgated as having purposes and uses including staff development, curriculum reform, diagnosis and reteaching of students, student certification, accountability, job selection, and college or other post-secondary admissions. This is a tall order.

Knowledge Base for Alternative Assessment

The research base on alternative or performance assessment has been described elsewhere (Baker, 1990) but, in sum, we know relatively little about the extent to which alternative assessment is successful in meeting the range of goals identified for its use. Three major sources of information are assessments in other countries, assessments in the military, and the field of writing assessment.

In brief, the evidence from the international community has only limited relevance in the United States context. First, no other country has the psychometric standards -- of validity, reliability, and fairness -- that are common in the United States. The guidelines, articulated by the Standards for Educational and Psychological Measurement, would not be met in any other country in the world. Part of the explanation for this is the psychometric perspective and expertise in this country. But another reason is the propensity of Americans to litigate on the grounds of fairness when test results are used for purposes with serious outcomes, i.e., high stakes, for an individual or system. Much of the technical quality concern in assessment is generated as either offensive or defensive measures from potential litigants in testing enterprises.

Although essay examinations are widespread internationally, with scoring schemes that range from explicit to imaginary, with very few exceptions, e.g., the Netherlands and Israel, school based assessment is focused on written performance. A recent national policy experiment in Great Britain promoted the use of hands-on alternative assessments in their school systems. The early results suggested that this process had many administrative and resource problems. Teachers were apparently unable to devote the specific, detailed attention needed to judge students' responses and simultaneously maintain the order and pace of instruction for those not being tested at the moment. As a result, there is a general regrouping and rethinking of the utility of this approach.

In the context of vocational training and testing, there are tests in Germany which require particular performance, occupation by occupation. These assessments are integrally linked with the apprentice and other training programs available to non-university bound youth. Studies of this system may be useful for future U.S. analysis.

A second source of information comes from a review of performance assessment in use in the military. Although job performance testing occurs in the assigned unit, the military, for reasons of cost, has stopped using some of the major performance testing, particularly the Skills Qualification test. Although considerable research has been conducted on performance assessments in the military, they have been generally focused on predicting proficient performance from other measures (see Wigdor and Greene, 1991). What is clear is that, with sufficient resources, large scale administration of performance assessments is possible. The military tasks, by and large, focus on identification and procedural tasks, and rarely deal with the conceptual, problem solving, or integration tasks that are the goals of more general educational programs. What is also clear is that such assessments can be subject to bias or corruption as well. When quotas are desired, performance ratings can be manipulated. This almost endemic effect of accountability testing is certainly not avoided because of the type of test used -- performance, multiple-choice, or otherwise.

Research on writing assessment provides the third sector from which we can draw inferences about performance or alternative assessments. Evidence suggests that raters can be reliably trained to make complex judgments, and that these judgments can adhere to an explicit set of criteria, rather than simply on judgments of good and poor performance. Raters can also be helped through specific procedures during the scoring process to cleave to the explicit criteria and not succumb to fatigue or socially redefined categories of judgment. These points are essential if one believes that the rating scale should have direct implication for the instructional activities.

Standards for Quality Alternative Assessments

One effort by CRESST has been to generate a first set of criteria to use in the evaluation of performance assessments. Part of these criteria are applied by inspection. One reviews the assessment and makes judgments about the extent to which it exemplifies the standards. These criteria include whether the assessment is meaningful to students and teachers; whether the content assessed is of high quality; whether there is adequate content coverage; and whether the assessment calls for complex cognitions on the part of the learning. External criteria include whether the assessment promotes generalization and transfer, its fairness, and its cost and administrative practicality. Most important is the consequences that using such an assessment has on the quality of learning and schooling, a dimension difficult to measure but one that should be kept in mind. Although these criteria come from many sources, including the writings of Messick and others, we believe that research studies can be operationalized to assess them as new performance assessments are designed.

Before alternative assessment should become a national policy, there are several areas of work to be done, work quite apart from technical standards.

Evidence of Impact

While there is almost astrological belief that improved assessments will magnetically pull teaching and learning into planetary alignment, what is the evidence for such expectations? Some argue that because multiple-choice tests negatively influenced teaching and led to adaptation to increase scores, e.g., training in test-wiseness and a molecularized curriculum, they believe that setting high standards for assessment will exert control on, of a more positive sort, the instructional behaviors of teachers. One commonly cited source of evidence for this assertion is performance in writing assessment. A particular example is the reputed impact of the implementation of the California Assessment Program (CAP) writing assessment. Data from San Diego School District suggest that writing performance has dramatically improved on most types of writing assessed by CAP over the last three years (Raines & Behnke, 1991). Yet, as the Raines and Behnke report suggests, considerable efforts in staff development were made in parallel to the advent of the CAP writing assessment. Furthermore, staff development did not have to start cold. In California, there has been a strong and continuing effort by virtually all major post-secondary colleges and universities to support improved instruction in writing through the California Writing Project. The conceptual and, to some extent, procedural analyses requisite for the design of staff development preceded the CAP writing assessment by at least a decade. How ready are disciplines other than writing to provide staff development with a coherent conceptual framework and valid delivery system?

Clarify What is Meant by Alternative Assessment

Enormous confusion and a lot of sloppiness exist in the use of terms. What are we talking about? Passion and description are intertwined. Authentic assessment is a case in point. The term connotes assessment "better than your kind," more real and deserving attention. In practice, it could be used to denote assessments that are more contextualized and either simulate or use performance derived from everyday, non-school tasks. Another inference for the term is that the assessment stimulates more genuine and representative samples of student work because it has more implicit meaning to them. This interpretation is rich in research opportunities. Alternative assessment means anything but multiple-choice (and problem true-false) but generally connotes extended and multi-step production tasks. Such tasks inevitably require the use of raters, judges, or their electronic proxies to determine the quality of the student's effort. Performance assessment encompasses both the meanings above and may specifically call up tasks that require either hands-on activity for solution or tasks where the student solution processes (in science) or ephemeral acts (speech-giving) must be observed.

Alternative assessment definitions must include the designation of the type of intellectual skill assessed (such as explanation or problem solving) and how they interact with sexy format changes. A portfolio is not a portfolio is not a portfolio. We need to hurry the process through while a generally agreed upon lexicon emerges.

Procedures for Developing Performance Assessment Need to be Clear and Consequences of Alternative Strategies Tested

Procedures for developing alternative assessments vary widely and are built mostly on trust. At the heart of the question of development are two issues: first, what is being assessed; second, how will the assessment be used? To the first point, if the assessment is to serve in any way as a standard to demonstrate competency for individuals or to provide a mark for system performance, the identification of the intellectual processes and content/situation domains must be identified. Assessments do not teach by themselves. How are teachers to know which types of instructional tasks are likely to prepare students for alternative assessments if the underpinnings of these assessments are not described in terms the teacher can understand. Some explication of the intention and class of performance of which the alternative is an example must be described. This stricture assumes that at least some alternative assessment attempts to provide a general framework in which to place students' accomplishments. Task specification seems an obvious option (Baker, Niemi, Aschbacher, Ni, & Yamaguchi, 1991).

The second issue, the purpose for the assessment, forces a consideration of the issue of the representativeness of student performance on alternative assessments. Given the extended time periods and resources used in many alternative assessment, we need to feel that our findings are trustworthy and fairly represent student capability. Research (Shavelson, 1990; Linn, 1991; Baker, et al., 1991), and pronouncements (Hoover, 1991), suggest that task sampling is a major validity issue. Specifically, researchers have found only moderate correlations between a given student's performance over a set of different tasks. This phenomenon may be due to lack of coherent specifications of the performance task domain, lack of coherent instructional experience, or the inherent instability of more complex performance. Recent research shows some prospect for controlling topic variability (Baker, 1992; Shavelson, Gao, & Baxter 1992) but until some replicated insight on this phenomenon can be developed, using performance assessments for individual student decisions is a scary prospect.

Format and Criteria: Two Critical Features of Alternative Assessment

Among practitioners there is a disconcerting tendency to overvalue differences in format, e.g., hands-on, portfolio, multi-step performance, and leave the identification of scoring criteria " 'til later." Alternative formats for performance are certainly the salient elements of performance assessment. The push for authenticity, that is, the context-sensitive nature of the assessment task, is supported by legions of research in cognitive psychology although this view shows some sign of revisionist thinking. Nonetheless, it simply does not make sense to generate tasks without knowing how or whether they can be credibly scored.

How should scoring rubrics be generated? The most frequent strategy seems to be assembling groups of teachers to decide on scoring dimensions. Evidence from our own research suggests that teachers are not good identifiers of criteria for certain aspects of student performance. For example, we found that teacher-generated criteria could not be transferred in training to other teachers. It was only after we analyzed performances of experts in contrast to teachers and students that we were able to develop scoring rubrics that teachers could be trained to use reliably and that showed desired relationships among other types of student performance and teachers' judgments. These criteria include the students' use of prior knowledge, principles, newly acquired information, and avoidance of misconceptions and, to date, they seem to work well in explanation tasks for history and science. Although we believe criteria should be generated or selected at the time the assessment task is developed, comparative research could be conducted on the cost, feasibility, and resulting quality assessments developed with different models.

In addition, within particular fields, such as writing or history, there are ideological differences of opinion regarding which set of criteria should be employed and whether, for instance, every new task requires its own specially crafted set of scoring criteria. Obviously, such issues are researchable, and a team of us are conducting studies assessing the robustness and validity of alternative kinds of scoring criteria.

The importance of identifiable and public criteria cannot be underestimated. Many analysts have distinguished between the need for common criteria for accountability purposes and the use of teachers' idiosyncratic criteria for assessment in their own classrooms. However, it is clear that equity concerns must drive us in the direction of having common understandings and standards for performance for both accountability and instructional purposes if performance disparities are to be reduced. Yet, if students in different schools are being held to vastly different types of performance, equity issues will exponentially increase with performance assessment.

Adult Views are not Student Views of Assessment

Much is made of the meaningfulness and challenge of alternative assessments as a means to renew students' interest and commitment to school. Our research suggests that students are not nearly so entranced as we are with challenging tests. There is evidence that students do not attempt tasks that seem long and hard. Our studies of anxiety show significant negative relationships with performance on alternative assessments and relatively high levels of anxiety. If students are not willing to engage in such tasks, then our efforts to estimate their performance will be thwarted. The lack of student interest may be a transitional problem, ameliorated following exposure to appropriate instruction.

Educational Equity

Alternative assessment will generate bad news in the short run. Our research in history and science show students have extremely low levels of understanding. Performance is low across the board--terrible for simple short answer assessment of knowledge, those elements of the curriculum thought to be supported by the use of multiple choice tests. Performance in complex explanation, for instance, integrating prior-knowledge with principle-driven explanation is lower still. Students don't know how to do what is expected of them in these tasks, and they report that they have not been taught such tasks in school. The dilemma is that we cannot improve the quality of these tasks, nor even understand much about their properties, until we can conduct research on students with more than a modicum of knowledge. We need to do teaching experiments to document the obvious proposition that instruction can impact alternative assessment performance. Teachers are going to need to be taught.

Massive support is needed to make alternative assessment a successful reform. Students don't perform well on alternative assessments because teachers have not taught them to do so. Many assume that teachers know how to teach complex cognitive skills but do not do so because of inhibiting multiple choice tests, unresponsive administrations, and so forth. I believe that people do what they know how to do. And I imagine that many teachers simply don't know how to approach instruction of the sort we are describing. We can explain their lack of expertise variously, but it is more important that we consider how to remedy it. For new forms of assessment to have a chance, enormous levels of staff development support must be available to practicing teachers. Significant aspects of teacher education programs must be seriously revamped. Such ambitions require resources. Many agencies are grappling with this problem. For example, the state of California is contemplating a major change in assessment and is exploring options to secure adequate support for staff development. Clearly, the state cannot simply down-load staff development responsibilities, including the continuing design and scoring of assessments, to local districts. We may have even a bigger problem, because redesigned staff development assumes we know what we want to teach teachers to do -- an unsupported proposition.

Beyond resources for assessment and staff, systematic development, implementing alternative assessment has additional costs. On the mundane level, teachers have told us they need additional teaching assistant time simply to use and to manage students during alternative assessments themselves, let alone change their teaching strategies. Costs for copying and materials will rise and this set of resource problems crops up just as local school districts are scaling back dramatically in the face of economic downturn and voters' reluctance to support additional costs for schools.

Equity issues are critical for alternative assessment. Equity has been at the heart of many advances in assessment and underscores some arguments against traditional testing (National Commission on Testing and Public Policy, 1990; Baker & Stites, 1991). Yet, almost paradoxically, the alternative assessment movement faces almost paralyzing equity challenges. First, there is a critical need to educate all but especially minority communities about new developments in assessment. This need is made more intensive by community suspicion that the establishment is once more changing the game and creating a new barrier by moving away from a known method of testing. Second, the very scoring of alternative assessments based, as they are, on students' observed performance (as opposed to products), raises equity concerns. Raters' (or teachers') expectations may be affected by race and ethnicity. Safeguards will need to be put in place and potential bias will need to be assessed and accounted for. Third, disadvantaged students may suffer disproportionately from their teachers' lack of experience in teaching complex tasks if for no other reason than these students will not so frequently be exposed to compensatory experiences in the home. One way to assist in reducing the disparities is to assure that students have been exposed to desired material. Although reports of simple exposure or opportunity to learn are pale reflections of whether students have had useful and sensible instruction, they are far better than nothing. In a state such as California, with a set of clear curriculum frameworks, classrooms can be monitored on their adherence to such blueprints (CAP, 1991). In fact, we have suggested using portfolios as an indicator of curriculum exposure rather than only or even as an outcome measure (Baker & Linn, 1990). Most importantly, reports of student performance should be conditioned by data on instructional exposure. Nonetheless, we can expect the gap between disadvantaged and economically secure students to widen dramatically. The only saving grace is that when the gap in their performance eventually narrows, the results should have deeper meaning. Evidence to date suggests that such gaps are present between certain ethnicities.

Educational Equity and a National System of Examinations

The report, Raising Standards for American Education, speaks to the equity concerns associated with any national system of assessment. The report recommends that no single test be used for any subject matter and grade level. It supports the development of local examinations to assess the national standards and specifies that a national quality control mechanism, consisting of a review board made up of experts, educators, and the public, oversee the quality of the measures.

This oversight is especially critical when any national examination is to be used for accountability purposes, for instance, to assess the quality of particular programs. A major precept, included in the Appendix to the report, specifies that states or clusters of states who wish their assessment reviewed must provide evidence of validity of the assessment for its purpose and equity interests. Specifically, the report says, The entity (quality control board) will design, in consultation with state and local educators, guidelines for the collection of evidence on system and school delivery indicators, with specific attention to equity protection. Decisions will be made related to the differential need for delivery indicators for different assessment purposes. States will provide such evidence as it becomes available. When evidence of both delivery indicators and validity standards is adequate, the entity will support the use of high-stakes assessment with secondary school students. It is anticipated that the entity will conduct audit studies, by visiting samples of schools, to verify the delivery and equity evidence provided by states.

States will (also) come forward with their plans for assuring equity in assessment design, administration, and use for gender, for special populations, disadvantaged students, and Limited English Proficient (LEP) students for review by this entity.

There are three principal concerns regarding equity in assessment of LEP and other student populations:

- If students are not assessed because of the lack of instruments, they will fail to benefit from the presumed desirable effects of assessment (improved instruction, accountability, and targeting of resources).
- If LEP students are assessed in English on subject matters such as mathematics, their performance will be handicapped to varying degrees by their English skills. The problem is not easily resolved even by assessment through the native language because of the heterogeneity of students and instructional programs for LEP students. Special procedures will need to be developed to take language and culture into consideration for appropriate assessment.
- All students must be provided opportunity to learn.

Conclusion

Because new forms of testing have a fragile research base, come at high cost, and present significant challenges to the educational community, we are going to have to use them wisely. Rhapsodizing on the wonders of these assessments makes no sense without thinking in parallel about real problems: about issues such as what and how information follows the student from grade to grade, school to school, or district to district; about how to get information on student content expertise, intellectual skill, motivation, and group cooperation all from the same assessment; about how technology can rapidly be employed to make sense of this process; about how we'll know we've been successful. Although many see alternative assessment predominantly in a personal, interactive, and dynamic classroom environment (Wolf, 1990), one challenge to smarter assessment is whether and how to project alternative assessment simultaneously onto the canvas of large scale assessment. Our interest is to design assessments to serve both instructional and accountability needs. We are unlikely to be successful completely but, for certain definitions of accountability, we probably can make progress (see Burstein, 1991) and justify the expenditure in this area. We have begun to design a theory of assessment that permits simultaneous information for both broad policy and teaching uses of assessment (Baker, Freeman, & Clayton, 1991). This parallel attention to policy and teaching purposes radically revises the common litany of assessment--that separate and different measures are always for different purposes.

Appendix

National Education Goals: By the Year 2000:

Goal 1:

Readiness for School: All children in America will start school ready to learn.

Goal 2:

High School Completion: High school graduation rate will increase to at least 90 percent.

Goal 3:

Student Achievement and Citizenship: American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.

Goal 4:

Science and Mathematics: U.S. students will be first in the world in science and mathematics achievement.

Appendix (Continued)

Goal 5:

Adult Literacy and Lifelong Learning: Every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.

Goal 6:

Safe, Disciplined, and Drug-Free Schools: Every school in America will be free of drugs and violence and will offer a disciplined environment conducive to learning.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Ambach, G. (1991, March). Improving curriculum and instruction. Paper presented at the meeting "Educational Assessment for the 21st Century: The National Agenda," of the Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

Baker, E.L. (1992). The role of domain specifications in improving the technical quality of performance assessment (CRESST deliverable). Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.

Baker, E.L., Aschbacher, P., Niemi, D., Chang, S.C., Weinstock, M., & Herl, H. (1991). Validating measures of deep understanding of history. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Baker, E.L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds.), Testing and cognition. Englewood Cliffs, NJ: Prentice-Hall.

Baker, E.L., Niemi, D., Aschbacher, P., Ni, Y., & Yamaguchi, E. (1991). Using cognitively sensitive assessments of history. In E. L. Baker (Ed.), Designing and scoring content assessments in American history. Los Angeles: UCLA Center for the Study of Evaluation.

Baker, E.L., & Stites, R. (1991). Trends in testing in the USA. In Politics of education association yearbook 90. London: Taylor & Francis.

Baker, E.L. (1990). Assessment and public policy: Does validity matter? Paper presented at the American Evaluation Association Annual Meeting, Washington, D.C.

Baker, E.L., & Linn, R.L. (1990). Advancing educational quality through learning-based assessment, evaluation, and testing (institutional grant proposal for OERI center on Assessment, Evaluation, and Testing). Los Angeles: UCLA Center for Research on Evaluation, Standards, and Student Testing.

Baron, J.B. (1990). How science is tested and taught in elementary school science classrooms. Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Burstein, L. (1991). Performance assessment for accountability purposes: Taking the plunge and assessing the consequences. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

California Assessment Program (1991). New Integrated Assessment System for California Schools. Los Angeles.

Hoover, H.D. (1991). Some cautions regarding the use of "alternative" assessments in high stakes situations. Presented at the UCLA/CRESST conference Educational Assessments for the Twenty-First Century: The National Agenda, Los Angeles.

Linn, R.L. (1991). Alternative forms of assessment: Implications for measurement. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

National Commission on Testing and Public Policy (1990). From gatekeeper to gateway: Transforming testing in America. Chestnut Hill, MA: Author.

The National Council on Education Standards and Testing (1992). Raising standards for American education. Washington: U.S. Government Printing Office.

Raines, R., & Behnke, G. (1991). California assessment program direct writing assessment statewide testing results by district and by school 1989-90. San Diego, CA: San Diego City Schools.

Resnick, L. (1990). Assessment and educational standards. Presentation to The Promise and Peril of alternative Assessment Conference, Washington.

Shavelson, R.J., Gao, X., & Baxter, G.P. (1992). Content validity of performance assessments: Centrality of domain specifications. Unpublished manuscript, University of California, Santa Barbara, CA.

Shavelson, R. (1990). Alternative technologies for assessing achievement. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Torney-Purta, J.V. (1990). Measurement of performance in social studies. Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Wigdor, A.K., & Green, B.F., Jr. (1991). Performance assessment in the work place, Volume I. Washington: National Academy Press.

Wolf, D.P. (1990). Assessment as an episode of learning: Making new assessments broad enough. Presentation to the Department of Education conference on alternative assessment.

498

Response to Eva Baker's Presentation

Lorraine Valdez Pierce
Center for Applied Linguistics

Let me begin by saying that I would like to commend Dr. Baker and her colleagues at CRESST for the impressive work being conducted at their Center. I am particularly impressed by the thoroughness of the studies on alternative assessment, even though they are still limited in number.

In her paper, Dr. Baker sets out four tasks for herself. These were:

(1) To describe and define alternative assessment and its characteristics and comment on these;

(2) To review the evidence in support of alternative assessment or performance-based assessment;

(3) To consider the validity of alternative assessment when it is applied under "various policy options;" and

(4) To present an example of research and development in alternative assessment being conducted at CRESST.

Dr. Baker accomplishes these tasks admirably and in her comments raises many problematic issues, among them, the need to define and clarify terms and purposes for alternative assessment and to ensure that alternative assessments are valid for the purposes for which they are used.

I found very little to disagree with in Dr. Baker's paper in regard to general definitions and uses of alternative assessment and problematic issues in ensuring validity. However, as a discussant at this symposium, I welcome the opportunity to expand upon some of the issues raised in the paper and suggest some things CRESST might consider examining in future studies. Because of the time limitations, I will not address what I see as less significant points I might tend to disagree with but, instead, focus on the key issues raised in the paper.

First, I will focus on purposes of alternative assessment, or how it is used. Second, I will address implications of alternative assessment and high-stakes testing for English language learners. Third, I will address the appropriateness and feasibility of using alternative assessment measures in high-stakes testing programs. For this, I will draw upon our experiences at the Evaluation Assistance Center-East at Georgetown University in assisting local and state education agencies to conceptualize, design, administer, score, and interpret alternative assessment instruments for students acquiring English, including the design of portfolio assessment systems. And fourth, I will propose recommendations for making future studies on alternative assessment more relevant to the linguistic, cultural, and academic needs of students learning English as their second language.

Purposes of Alternative Assessment

Dr. Baker proposes at least six different purposes of alternative assessment and states that these purposes differ in relation to the broader policy context and to the technical demands they place on the quality of the assessment. The purposes she names include what we can refer to as low-stakes uses, such as school reform, instructional improvement, and grading, and high-stakes uses, such as certification (national testing) and selection (college admission). In a key piece, she states that by wanting to combine both low and high-stakes purposes, practitioners tend to confound problematic issues in validity because they confuse the purposes of alternative assessment with those of traditional, student achievement testing.

While I tend to lean in favor of this argument, I also sense some dissension in the field with regard to the purposes or uses to which alternative assessment can be put. Controversy tends to be inevitable when educational innovations are under consideration. On the one hand, in response to increasing dissatisfaction with the limited information provided by multiple-choice achievement test formats, especially with regard to students not yet proficient in English, practitioners are turning to alternative assessment as a tool not only for identification of students of limited English proficiency but also for monitoring student progress on a continuous and frequent basis. One of the advantages which practitioners perceive alternative assessment to have over once-a-year standardized achievement testing is its potential for providing multiple sources of information over time on student progress in language proficiency and content area knowledge.

On the other hand, we know of states that have in the past or are presently attempting to incorporate alternative assessment measures, such as writing samples, into their statewide testing programs. Some of the states that have been using alternative assessment are in our immediate area, such as Virginia, (in the form of the Literacy Passport Test) and Maryland (in its Functional Literacy Test). Both of these states include student writing samples as part of their statewide testing program. The state of Michigan has also determined alternative assessment to be not only possible but feasible for large-scale, high-stakes purposes. New York has demonstrated that it is feasible to administer performance tests to every pupil in science. This year Connecticut has implemented the first statewide portfolio assessment system in the nation. By applying alternative assessment in their high-stakes testing programs, these states and others are telling the rest of the nation that they are willing to use alternative assessment and performance assessment to determine whether or not students have achieved the skills that the states most want them to learn. These skills include synthesis and application of individual bits of knowledge. States may be using alternative assessment because they have found that whereas students may do well on multiple-choice tests, this is insufficient evidence for concluding that they can also integrate these facts and skills into desired performance outcomes.

What I suggest we need to look at is how states and local school districts are using the data resulting from the use of alternative assessment measures in large-scale testing programs. I think we can safely assume that they are using the results for the same purposes for which they have used traditional, standardized achievement test results. The following five questions come to mind:

(1) Are states using the data to compare students in order to determine program effectiveness? If they are, students may not have received equal opportunities to learn, may not have participated in the same programs, or may be limited in their English proficiency.

(2) Are states and/or school districts using the results to meet grade promotion and graduation requirements? The research indicates that grade retention is not an effective educational practice, especially with minority students.

(3) Are states using the results to track students? Equity issues indicate that tracking is unacceptable and illegal if ethnic/racial tracking results from these practices.

(4) Are states using the results of alternative assessment measures to provide special instructional services to students who did not attain the minimum score? and

(5) Are states using alternative assessment procedures providing guidelines for the participation or non-participation of LEP students in these statewide testing programs?

Depending on the answers to these questions, the validity of the alternative assessment results and the purposes to which these measures are put become terribly important. Yes, alternative assessment will continue to be used in large-scale programs and perhaps in a national assessment system, although, as Dr. Baker has suggested, this may lead to a headlong rush to use alternative assessment measures which may not be valid for the purpose for which they were designed.

Implications for Students Learning English

When we consider the implications of high-stakes testing using alternative assessment measures for the general student population, test reliability and validity become essential. But when we consider the implications of using these same alternative assessment instruments with students who are not yet fully proficient in English, the reliability and validity of alternative assessment measures become critical. Although no reference is made in her paper to language minority students or to limited English proficient students in particular, I think some of the points made by Dr. Baker can be expanded upon in order to more clearly see the implications for these students.

First, the review of the literature conducted by Dr. Baker revealed data on the generalizability of performance across tasks. These data indicated that variations in task performance seem to be attributable, in addition to the degree to which tasks were comparable, to differences in specific prior knowledge, including the type of instruction received by students. The State-NAEP data seem to indicate that lower student performance in performance assessments may be a result of a lack of appropriate instructional experience. Students differed in the rate at which they attempted the more open-ended types of items. The implication is that students in "disadvantaged classrooms" who were not exposed to instructional experiences demanding complex performance were not as prepared to take the tests as those who were.

In the extensive studies conducted at CRESST on creating valid alternative assessment measures in the content areas, it was also determined that students brought a relatively low level of prior knowledge to the tasks and so performed poorly. In addition, it was noted that the researchers were "concerned about the heavy verbal load these tasks place on students."

I believe the implications for language minority students not yet proficient in English are clear: In addition to a possible lack of prior knowledge in the form of educational experiences and opportunities, the limited English proficient student also brings a lack of English language skills, including knowledge of the culture in many cases. Many of these students are placed in classrooms where complex performance is not expected and alternative assessment techniques are not used, taught, or practiced. Lacking this exposure, students in the process of acquiring English, who are required to take high-stakes tests that employ alternative assessment measures, are put at an additional disadvantage.

When we consider that students not yet proficient in English may be retained in grade or denied a high school diploma as a result of their performance on high-stakes alternative assessment measures, it becomes of paramount importance to either ensure that these students obtain access to the same kinds of instructional experiences that fluent English-speaking grademates have or that they obtain exemptions, waivers, or other considerations due to their language status, such as alternative assessment in the native language.

Feasibility of Alternative Assessment in Large-Scale Testing Programs

I share Dr. Baker's concern for valid alternative assessment and performance-based assessments and agree that such measures require time, conceptual models, and empirical studies to support their validity. However, I do not believe that performance measures are entirely inappropriate for large-scale assessments, given the following conditions:

(1) The purpose of the assessment is clear and the instrument has construct validity;

(2) Steps are taken to reduce cultural bias so that students from linguistically and culturally diverse backgrounds are not unnecessarily penalized;

(3) Procedures are specified for designing, administering, scoring, and interpreting each measure;

(4) Raters are trained in scoring procedures and inter-rater reliability is consistently high; and

(5) Results obtained on alternative assessment measures and traditional standardized achievement tests are used in combination as opposed to using a score from only one type of test or the other.

At the Georgetown University Evaluation Assistance Center-East, we have received increasing numbers of requests for technical assistance on alternative assessment and portfolio design. We have presented workshops on these topics to teachers and administrators in states all over the Eastern half of the United States, including Puerto Rico and the Virgin Islands, as well as at regional and national conferences. We believe that once practitioners are trained in how to determine the focus of the alternative assessment and in how to score and interpret the data, the potential for increasing the validity of the alternative assessment measure increases. We have suggested that portfolio planning committees composed of teachers and other staff clearly define the purposes of their assessment, select alternative assessment measures which they believe will match their purpose, and identify specific procedures and criteria for scoring and interpreting these measures. These committees also need to consider assigning weights to the relative value of each measure in a student portfolio. When we consider that teachers and administrators are designing the instruments and setting the standards, we can see that alternative assessment lends itself remarkably well to the setting of local accountability standards.

There is little reason to believe that, given the above-mentioned conditions, alternative assessment could not be implemented in large-scale assessment programs. But these are formidable conditions, similar to those specified by Dr. Baker in her discussion on methods for addressing the comparability of alternative assessments. As Dr. Baker notes, research on alternative assessment is still in its infancy, and we have a long way to go before its applicability to high-stakes testing is clear. In light of this, I would like to make some recommendations for ensuring that future research on alternative assessment addresses the needs of language minority students learning English.

Recommendations

Future studies on alternative assessment need to describe not only the purposes of alternative assessment measures and steps taken to ensure their validity but also key characteristics of the students involved in these studies, because not all language minority students are the same. Specifically, studies need to look at:

(1) Language minority students who are not limited in their English proficiency, representing all grades and skill levels;

(2) Students who are learning English across all grades and of varying levels of English language proficiency;

(3) Variations in students' prior educational background and literacy skills;

(4) The types of instructional programs in which students have participated;

(5) The effects of practicing alternative assessment techniques with language minority students having different levels of English language proficiency, from varying educational backgrounds, and who have participated in mainstream and special instructional programs;

(6) The purposes for which the alternative assessment is being conducted, whether for identification, entry into language support programs such as ESL or bilingual education, monitoring student progress, or exiting from language support programs into mainstream classrooms;

(7) The academic language skills needed for success in English language content-area classrooms, such as math and science, and developing alternative measures to assess the development of these skills for students for whom English is a second or additional language;

(8) The collaborative frameworks, such as school-wide portfolio assessment teams, which facilitate exchange of information between ESL, bilingual education, and mainstream teachers on portfolio assessment; and

(9) Innovative, informative staff development programs which enable teachers and school staff to use alternative assessment frequently and well. By taking into consideration student and instructional characteristics and the purposes of the assessment, we can better determine the potential of each alternative assessment measure to meet its purpose.

Response to Eva Baker's Presentation

Peter M. Byron
New York State Education Department

I'm sure all of you would agree that the previous speakers make my task very difficult. I have two very excellent acts to follow and only hope that my fifteen minute commentary will provide a useful addition to what you have previously heard.

It is an honor to comment on Dr. Eva Baker's paper. Regretfully, you have not had an opportunity to read her excellent paper. Take our word that this paper is well worth your time!

However, before sharing my thoughts on the paper, I would like to congratulate the Office of Bilingual Education and Minority Languages Affairs (OBEMLA) at the Department of Education for convening this symposium. As you are aware, alternative assessment is a topic currently under debate in the measurement community. OBEMLA provided us the service of directing our focus to a current measurement concern. The topic is extremely important because this measurement issue has not been limited to the education community. The issues surrounding assessment have become a part of our evening newscasts. Last night's news provided an endorsement of alternative assessment by Tom Brokaw and a photo opportunity for the President as he visited elementary and secondary schools in Maine. National education policy and assessment has moved to the front burner. The Office of Bilingual Education and Language Minority Affairs provided the field with an opportunity to become engaged in the national debate on education policy. Whether symposium participants agree with the testing instruments proposed or with the nature of the outcomes to be measured, each participant has been given a unique opportunity to listen and to engage in discussion of the national policy and return home better prepared to take an active part in the debate which will certainly follow at the local level.

The third service which OBEMLA provided is that the office combined a measurement community issue and a national policy debate and focused these on limited English proficient children. It is important that limited English proficient children be a part of the discussion from the very beginning and not introduced as an afterthought when policy decisions have been made. The assessment implications for limited English proficient children must be discussed at the very beginning and not when assessment practices are in place.

The Role of Identification, Assessment, and Evaluation

Those with a traditional background in bilingual education recognize that most decisions and most programs depend on three very important issues: identification, assessment, and evaluation.

Dr. Baker and Dr. Valdez Pierce spoke of high-stakes testing. All testing with limited English proficient students is high-stakes. Educators of limited English proficient students must realize that whenever testing is discussed, they must listen. Court decisions, state programs and federal programs all depend on testing and ultimately identification, assessment, and evaluation.

When we fail to appropriately identify limited English proficient students, we lose them. The debate about the number of limited English proficient students in this country is not purely academic. Programs are designed on needs and programs will not be designed if the needs are not properly identified. Identification is essential to program planning.

However, once limited English proficient students are identified they must be placed in appropriate programs. Without valid and reliable assessment or placement practices, students are placed in programs which are not designed for their needs. Whereas identification and assessment are important, program evaluation is essential. Without program evaluation standards, we cannot begin to tell the story about how successful our programs are, and we certainly can't modify programs, if modification is needed.

The discussion we have today and our continuing discussion over the next two or three days is extremely important if not critical for us because future services for limited English proficient students will depend on our deliberations. I ask that each of you keep the words identification, assessment, and evaluation in mind as presenters discuss portfolio assessment, assessment in science and assessment in mathematics and question whether these procedures will result in fair, equitable, and appropriate treatment for limited English proficient students. We must focus our attention on the students. Remember, without appropriate identification there are no students. Without appropriate assessment, the students are placed in the wrong programs and without appropriate evaluation, we can't tell the story of what we do or what the students accomplish. Let's examine why alternative assessment is important by reviewing Dr. Baker's paper.

Alternative Assessment

What is alternative assessment? The literature describes alternative assessment as an alternative to standardized testing which arose because opponents thought that among other problems, standardized tests: (1) provided false student information, (2) were biased against certain students, (3) focused on lower level skills, and (4) allowed teachers to teach to the test.

In commenting on Dr. Baker's paper, I will focus first on her style and then on the content of her paper and conclude with the message which I received from her work.

Style

Dr. Baker is at once crisp, concise, frugal and economical in her use of the English language. Her points are made without redundancy and with a certain imagery that is lacking in many research articles of this genre. I sensed from the paper that the author enjoys playing with words and invite each of you to take the time to test my hypothesis when the paper becomes available!

Content

How is the paper written? The author intended to explain the attributes of alternative assessment and provide examples of each. She did an excellent job. Dr. Baker is fair because she provided the proponent's position yet she is inquisitive in that she does not unquestioningly accept the proponent's view. An interesting aside is that Dr. Baker believes that many of the problems of standardized testing are also problems of alternative testing. Even though alternative assessment proponents talk about measuring higher order thinking skills, Dr. Baker notes that they may be focusing on simple order skills and many of the proponents also teach to the test, albeit a different type of test. She provides arguments from research which would call into question some assumptions made by the proponents.

Most research reviews conclude after criticisms, however, Dr. Baker provides a practitioner's perspective on alternative assessment from her role as a test developer who has operationalized what to others is only theory. The paper provides a primer on alternative assessment which would be difficult to match. The sole limitation which was mentioned by Dr. Valdez Pierce is the limited specific focus on limited English proficient students.

Message

What is the message? Although each of us brings a different perspective and receives a different message, I share these thoughts for your consideration when you read the paper.

This first message is that alternative assessment has a lot to offer. Alternative assessment has forced measurement specialists to rethink traditional practices and in this has contributed greatly to the field. However, hoping for a test and closing one's eyes and crossing one's fingers, does not make a test appear. Even though alternative assessment has a lot to offer, there is a long way to go before alternative assessment can be a reality.

The second message is that tests must successfully undergo a rigorous review by technical standards before they are deemed acceptable. The development of a test is not complete until it undergoes this review. Alternative assessment instruments must be held to technical standards. It may be that alternative assessment instruments and procedures may be more appropriate in areas other than high-stakes testing of limited English proficient students where identification and placement decisions which impact on a student's life choices are made. Alternative assessment may be more at home as a tool for classroom assessment.

In conclusion, because it appears my time is almost exhausted, I would like to shift focus and talk about future considerations. It would be shortsighted if this discussion were finished at the end of this symposium. For that reason, I propose:

1. A task force be convened by the Office of Bilingual Education and Minority Languages Affairs to discuss the testing practices and procedures used to identify and place limited English proficient students and evaluate educational programs serving them. This task force would be particularly charged to ensure that testing and evaluation standards be implemented which result in fair and equitable testing of limited English proficient students and that the testing of limited English proficient children will be a consideration in the development of all national educational programs.

2. Guidelines be developed to eliminate the unfair use of standardized testing in categorizing school populations. Item and population sampling may be investigated as possible interim solutions. This topic is one which could be the focus of another symposium.

3. Guidelines be developed to expand the use of alternative assessment practices in conjunction with standardized testing in the

program evaluation of federal education programs such as ESEA Title VII.

4. Incentives should be given for the development of computer technology for simulation testing which is an integral part of alternative methodologies.

My recommendations were addressed to the immediate concerns raised in this paper, however, it is important to realize that if we are to improve assessment practices for limited English proficient students we must begin simultaneous efforts in developing instrumentation in the child's native language, a topic which should also be the focus of a future symposium.

Testing Limited English Proficient Students for Minimum Competency and High School Graduation[1]

Kurt F. Geisinger
Fordham University

The Current Status of Minimum Competency Testing

Twenty years ago, only a handful of states in our union required students to pass a statewide examination to receive their high school diploma. Today, statewide high school competency tests are in use, in one form or another, in at least 40 states (Jaeger, 1989; Roeber, 1990). From an historical perspective, the widespread use of such examinations and assessments probably grew out of the "back to basics" movement which emerged in response to charges that many of the graduates of our educational system lacked the fundamental academic skills of reading, writing, and mathematics necessary to succeed in adult life, to hold useful and meaningful jobs, and to serve as responsible citizens. From the more limited psychometric or educational testing perspective, such tests probably developed out of the "criterion-referenced testing" movement which occurred in the period from approximately the mid-1960s through the early 1980s. The purpose of such tests was to integrate educational tests more meaningfully into the instructional process by reflecting exactly what knowledge, skills, and other educational behaviors students "mastered" and on which they therefore needed no further instruction. Criterion-referenced tests (CRTs) emphasize scores relevant to the knowledge domain and strongly de-emphasize comparisons of individual students with other children composing their norm group.

To combat charges that students were graduating from high school without being able to read, states imposed tests that students would need to pass to earn their high school diplomas, regardless of how well they had performed (e.g., in terms of grades) in their educational course work. In some cases, the tests were mandated by a state's department of education, or a similar body responsible for monitoring education within its jurisdiction. In other instances, the state legislature imposed the testing program on the educational community. Such tests then serve as a guarantee to society at large, parents, schoolchildren, potential employers, and others that high school graduates possess at least those minimal skills usually deemed necessary for successful survival in the modern world. Clearly, the responsibility for ensuring that those educated within a given school district or state falls on both those charged with monitoring education within the jurisdiction and those with overall responsibility for governing the region.

Although the use of minimum competency tests for scrutinizing whether students have acceptable levels of the skills measured by the tests to be promoted or graduated is the most visible use of these tests, it should be noted that there are other uses to which the scores may be put as well. For example, Roeber (1990) provides a number of additional uses: appraising the general mastery by students of the state curriculum (for program evaluation uses); providing general information to policy makers, educators, or the public; system accountability; system planning and resource allocation; and system improvement. In addition, when administered prior to the terminal year of a student's education, the test may be used to help to direct a student to a particular school stream (e.g., vocational or academic system) (Roeber, 1990, p. 7-8).

For those states in which passing the minimum competency test was required for high school graduation, states frequently phased in their offering of minimum competency examinations by one of several methods: by administering them to the first one or more classes on a trial basis, by permitting students to take the tests on several occasions in order to pass, by gradually raising an initially lower standard of performance until the appropriate, planned passing score is achieved, or some combination of these techniques. Students need only pass the test once; there is no requirement that they demonstrate continued competency once they have passed the examination.

Most states that began offering such tests did so by administering them at least a year or more before the expected date of graduation, so that students who failed to pass them would be able to retake the tests on one or more future occasions in order to graduate from high school on schedule. The Director of Testing for one state, which employs a high school graduation test, recently stated that students who repeatedly fail the test and take it each time it is offered could conceivably take it as many as 11 times. He further informed me that some students had indeed taken the test this number of times. In other states, of course, the number of possible retestings is considerably reduced. In North Carolina, for example, students must pass both reading and mathematics tests; they are given a maximum of four trials to pass each (Jaeger, 1989). Students who never pass but choose to leave high school typically receive a certificate of completion or some similar acknowledgment that the student completed his or her studies but did not graduate.

In many states, simply passing the competency tests does not ensure that a student will receive a high school diploma. Rather, it is one requirement among several, such as satisfaction of attendance policy, curricular breadth, and quality of academic course work requirements.

Many states that administer minimum competency tests also offer other preliminary tests at earlier points in the movement of students through the educational system. The purposes of these examinations is to identify those students who have fallen behind and who are likely to have problems when they eventually take the statewide graduation test. Students who perform poorly at these earlier grades may receive additional instruction, embellished instruction, and other special remedial services or be held in grade until they pass this preliminary examination.

How different states use minimum competency tests varies widely (Jaeger, 1989). The majority of states set standards which all students throughout the state must pass to earn their high school diploma. Others, however, permit each school district to set individual standards specific to their own school district. Still others do not require a passing score for high school graduation.

The passing of P.L. 94-142, the Education of All Handicapped Children Act, in 1975 mandated that all children be provided with a free and appropriate education, regardless of handicap. After this law went into effect, most states had to accommodate students with handicapping conditions and other disadvantaged students differentially. The law required the development of Individualized Educational Programs (IEPs) for all students with handicapping conditions. Description of the following items were mandated for inclusion in IEPs: a statement of the present levels of educational performance, short-term and long-term goals and objectives, specific educational services to be provided, the extent to which the student should participate in regular educational programs, the projected date for initiation of remedial services, the duration of the remedial services, and "appropriate objective criteria and evaluation procedures and schedules for determining, on at least an annual basis, whether instructional objectives are being achieved" (Willig & Ortiz, 1991, p. 282). Students with handicapping conditions, on the basis of their IEPs, can be rightfully exempted from the requirements to take and to pass the minimum competency examination in order to graduate from high school. However, unless LEP students also fit the criteria for handicapped status, no IEPs are developed for them. Because of this exclusion, "educators frequently fail to consider cultural/linguistic learner characteristics and their effects on the teaching-learning process" (Willig & Ortiz, 1991, p. 282). Thus, although some LEP children may be considered exceptional and others reside in states which provide special status to LEP students, most still need to pass statewide competency tests.

The Typical Content of State Minimum Competency Tests

A 1979 review (Gorth & Perkins, 1979) summarized the content of statewide competency examinations. (This information is summarized by Jaeger, 1989.) In general, two overlapping types of content are called for by these examinations: the basics of education (reading, writing, and arithmetic) and what are sometimes called "survival" skills for adults in our society. There is, of course, much overlap between the two content areas. Gorth and Perkins reported that over one-half of the states then using such examinations employed tests composed of multiple-choice questions of reading, writing, and arithmetic. In general, these examinations called for the students to demonstrate "nothing more than recognition of basic subject matter mechanics or the application of basic mechanics to so-called 'life skills' situations" (Jaeger, 1989, p. 510). Indeed, the tests were seen as measuring skills learned primarily at the elementary school level rather than either those drawing upon the high school curriculum or higher-order thinking processes.

Increasingly, states and district-level minimum competency examinations are including performance assessment components as parts of their competency testing program in addition to the traditional objective, multiple-choice test components. Such assessments are seen as differing from multiple-choice testing in that (1) students create responses rather than selecting them, (2) performance assessments emphasize problem solving and other higher-level integrative cognitive skills, and (3) performance assessments need to be scored by expert judges rather than machines (Finch, 1991). Because the skills that students use in generating their responses and the products that result from their responses are sometimes seen as more like those skills and products found in the classroom, performance assessment has occasionally been called authentic assessment. Among the types of performance assessment that are used are essays, sometimes with prompts provided; actual student writing samples; prepared portfolios which document the accumulated work of a student; problem solutions such as lab reports in the sciences; and reviews of productions in the realm of art and music. The most commonly used performance assessment component is the writing sample or essay as a measure of student writing ability (Roeber, 1990). These are sometimes administered as part of the examination process and in other settings students may write their essays during a time period of several weeks. A number of states are currently making efforts to increase their utilization of this form of performance assessment, especially in math and the sciences. The development and scoring of performance assessment measures is an extremely expensive undertaking. Therefore, performance assessment is likely to remain a component of minimum competency testing in conjunction with objective measurement (e.g., multiple-choice tests) and/or as an alternative assessment device for those individuals who fail the objective test on one or more occasions.

The American Achievement Tests called for by the Department of Education in the AMERICA 2000 report (U. S. Department of Education, 1991) would seem to draw upon similar skills, although they would appear to be both heightened in terms of difficulty and level of cognitive processing and broadened in scope. Five subject matter areas will be addressed (English, mathematics, science, history and geography), although when the tests are first introduced, they may be limited to an assessment of reading, writing and arithmetic. The tests would appear to be conceived as both tied to subject matter and to broader thinking skills as are more typically found in tests of cognitive abilities than are subject-area tests. Like other competency examinations, preliminary competency tests will be administered at earlier grades. Thus, "American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy" (U.S. Department of Education, 1991, p. 9). Frankly, one might legitimately question whether modern psychometrics, educational testing, and psychology have advanced to the stage of being able to identify those skills necessary for responsible citizenship, much less to measure them. It may be noted, however, that the American Achievement Tests called for in the AMERICA 2000 report are not necessarily conceived of as minimum competency tests that would be used for high school graduation. It is also suggested that they be used for college admissions and employment decision making, for example. How a single test could meet these varying purposes is not clear and difficult to imagine from the current status of test construction and theory.

Assessing LEP Students with Minimum Competency Examinations

At present, states have no consistent manner in which LEP students are assessed on statewide or district-level minimum competency examinations. In some states, LEP students need to take the same minimum competency examinations under the same rules as other students to graduate or be promoted. In other jurisdictions, however, exceptional LEP students and those residing in locales that require individualized educational programs for LEP students may be exempted from the examination if their IEPs do not require them to take the examination to graduate from high school. (Such a plan is similar to a common approach for waiving this requirement for special education students.) In yet other locations, LEP students may be permitted to take the examination in their native language, or at least in some common languages, if they enter the American educational system late in their formal education. In some schools, when students fail one competency test, they are given the option to take an alternative measure, perhaps a performance assessment or a test in their native language. In still other settings, they may only take the examination if they have first failed the examination in English. Furthermore, these options simply sample some of the possibilities. Thus, there is a wide variety of choices from which the educational community may select in deciding how LEP students should be tested with minimum competency examinations.

A few examples may demonstrate the diversity of options available. In Connecticut, LEP students are required to be tested unless a planning and placement team decision rules otherwise. In Florida, LEP students are exempt from taking the graduation test during their first two years in an English-speaking school, but are still required to pass the graduation test to qualify for a regular diploma. Similarly, in Michigan, non-English speaking students enrolled in schools in the United States less than two years may be excluded from taking the tests. In Ohio, students may defer taking the test, but may not earn a diploma without passing it. In Georgia, English-as-Second Language (ESL) students take the tests unless the school and parent(s) or guardian agree it is not in the best interest of the student to take it in its current administration. In Maryland, LEP students must pass all four of the examinations that they offer to earn a high school diploma.[2]

Little guidance is available from the educational research literature regarding which of the possible approaches to testing LEP students would be preferred. In fact, a computerized literature review of the ERIC database using either competency and minimum competency examinations and limited English proficiency students as key words or descriptors yielded no references. Certainly, survey efforts similar to those routinely performed by national organizations (e.g., Roeber, 1990) to document the policies each state follows with regard to LEP students would be a most helpful first step. Full-blown comparative evaluation studies such as those performed on the Headstart evaluation contrasting the effectiveness of different strategies in working with LEP students are needed.

Methodological Issues in Minimum Competency Testing

Validity

Validation has been thoroughly described by Messick (1989), Anastasi (1988), and others and need receive only a cursory treatment here. Validation refers to the process of documenting that a test is being used in a justifiable fashion, typically as determined by research studies providing documented evidence supportive of its planned use. Cronbach (1971) has sometimes been credited with the notion that we do not validate tests, rather we validate the accuracy of inferences that we make from test scores. There are generally three acknowledged models or approaches to test validation: criterion-related, construct-related, and content-related. With regard to construct validation, Anastasi (1988) has written:

> The construct-related validity of a test is the extent to which the test may be said to measure a theoretical construct or trait....It derives from established interrelationships among behavioral measures. Construct-related validation requires the gradual accumulation of information from a variety of sources. Any data throwing light on the nature of the trait under consideration and the conditions affecting its development and manifestations represent appropriate evidence for this validation (p. 153; also cited in Geisinger, in press b).

"Criterion-related validity is based on the degree of empirical relationship, usually in terms of correlations or regressions, between the test scores and criterion scores" (Messick, 1989, p. 17). What differentiates orthodox criterion-related validation from the other typically empirical method of validation, (i.e., construct validation), is that criterion-related validation focuses upon "selected relationships with measures that are critical for a particular applied purpose in a specific applied setting" (Messick, 1989, p. 17). The basis for making assessments regarding content validation is "professional judgments about the relevance of the test content to the content of a particular behavioral domain of interest and about the representativeness with which item or task content covers that domain" (Messick, 1989, p. 17).

In 1974, Standards for Educational and Psychological Measures recommended for the first time that social consequences of testing such as adverse impact and test bias, should be considered in evaluating a test (Geisinger, in press b). Indeed, Messick (e.g., 1975, 1980, 1989) has argued that values should guide both test use and test evaluation and, hence, such factors need to be considered in evaluating the use of tests and other measurement procedures. The use of tests with groups underrepresented in many settings within our society, such as LEP students, clearly invokes the values component of the evaluation of these measures.

Since minimum competency tests are the present focus and as educational tests represent a domain (of basic educational skills such as reading, writing, and arithmetic), content validation is the strategy most commonly associated with the corroboration of their use.

Content validation. The basis for content validation is "professional judgments about the relevance of the test content to the content of a particular behavioral domain of interest and about the representativeness with which item or task content covers that domain" (Messick, 1989, p. 17). Educational measures developed and validated using content validation involve carefully developed domain specifications based upon curricula, studies of actual instruction provided and educational goals. Content validation comprises both the relevance of the content called for by the test domain (or plan) as well as judgments regarding how well the test ultimately represents the test plan or domain.

Standard 8.4 of the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1985) deals with competency tests and is found below.

> When a test is to be used to certify the successful completion of a given level of education, either grade-to-grade promotion or high school graduation, both the test domain and the instructional domain at the given level of education should be described in sufficient detail, without compromising test security, so that the agreement between the test domain and the content domain can be evaluated (p. 54).

Because a test could meet the traditional standards of content validity yet fail to meet the criteria specified in Standard 8.4 above, demonstrates that traditional approaches to content validation do not provide the specificity called for by Standard 8.4. New concepts were required to gauge meaningfully the value of a minimum competency examination. These terms are curricular validity and instructional validity.

Curricular validity. The notion of curricular validity was introduced by McClung (1978, 1979). It is not a traditional type of validation called for by professional standards (such as content validation), but it has nevertheless, become an important principle in the evaluation of minimum competency examinations, especially in court cases (e.g., *Debra P. v. Turlington*, 1979, 1981, 1983, 1984). "Curricular validity is a measure of how well test items represent the objectives of the curriculum. An analysis of curricular validity would require comparison of the test objectives with the school's course objectives" (McClung, 1979, p. 682).

Instructional validity. A second characteristic that should be present in competency tests has been called instructional validity. "Instructional validity is an actual measure of whether the schools are providing students with instruction in the knowledge and skills required by the test" (McClung, 1979, p. 683). An assessment of instructional validity would require proof that students are actually exposed pedagogically to the content covered on the examination. Assuming that a state's minimum competency examination is valid for the majority of students in a state, an important question when considering the testing of LEP students is whether their instruction parallels that of the majority students. The concept of differential validation impacts this judgment.

Differential validity or population validity. Differential validity (sometimes referred to as population validity) is a concept closely aligned with that of test bias. It has traditionally been used in criterion-related validation studies, primarily with regard to admission to higher education and employment decisions. A test may be said to be differentially valid if its validity differs across subgroups of test takers. Predictive tests are differentially valid if the empirical relationships between the predictive test and a measure of criterion performance differ systematically across groups. To assess whether a test is differentially valid across groups, one must perform regressions between test and criterion variables for each population. Then, the slopes of the varying regression lines, their respective intercepts, and the degree to which the relationships are free from statistical error are compared. (See Anastasi, 1988, pp. 193-199 for an elaboration of this concept.) "Validity coefficients, regression weights, and cutoff scores may vary as a function of differences in the test takers' experiential backgrounds" (Anastasi, 1988, p. 194).

It is not clear how or why differential validation studies would be performed using minimum competency examinations since there is no criterion representing the kind of behavior that minimum competency tests attempt to measure or predict. In the present instance, however, there is no obvious, relevant criterion for a test of minimum competency. Possible criteria include teacher judgments, high school grade-point average (GPA), subsequent college GPA, etc., but all of these criteria have methodological problems and, more importantly, they lack relevancy in that none of these criteria bear on the meaning and purpose of such tests.

The concept of differential validity can nevertheless be generalized to the testing of LEP students. A competency test might be differentially valid in terms of *instructional validity* if the material covered on the examination is not equivalently presented to the majority students and LEP students. That is, if the material composing the examination is more to be found in the classroom of traditional students than it is in bilingual classrooms, a case could be made that the test is differentially valid and, in this instance, biased against the LEP students.

Reliability

The study of test reliability has largely been the study of the consistency of the test scores that individuals achieve across different administrations of the same test, across different test forms, across different test administrators (especially for individually administered examinations), and across the individual questions composing a single test. Each of these kinds of reliability indicates a somewhat different generalization about which we may have a degree of confidence when we talk about an individual's test score. Such depictions of test reliability do indeed have relevance for competency testing, but the relevance needs to be reformulated to a degree. The stability and consistency of an individual's score are important, but the degree to which the decisions made with the examination do not change is even more critical. These approaches are known as the decision-consistency approaches to test reliability. Excellent reviews of the literature on the reliability of tests scored in a pass-fail manner may be found in Berk (1984), Brennan (1984) and Subkoviak (1984).

The notion of stability over different testings must be clarified with regard to competency testing. Classical reliability theory, upon which the notion of both the test-retest and alternate-forms approaches to test reliability, are based, assumes no change in the underlying variable being measured. In the case of a competency test, however, it is certainly hoped and expected that instruction -- remedial and traditional -- increases the competency of the student between the first and second testings. Thus, the assumptions of the classical reliability model are clearly violated in the case of minimum competency testing. Indices lower than those that might otherwise be acceptable may be tolerated due to these expected changes over time. That is, since students are engaged in learning in their educational activities, their performance on the minimum competency tests changes from the first administration until the second. When this learning process occurs, it appears as though the test is less reliable (stable) than it really is.

Standards of Performance

Techniques used in setting standards. All certification testing requires that the performance of individual students be compared with or evaluated against a predetermined standard of performance.

A decision is made regarding each student in terms of whether that student is competent. The degree of competence is not critical as it is in most tests of individual differences or so-called norm-referenced tests. The only scoring that is critical is whether the student has met the minimal standard or not.

While psychometricians have developed theories for test reliability and validity, the development of such approaches for the setting of standards on examinations is in its infancy. A number of techniques or strategies for setting standards on educational and other psychological measures have been proposed, but it is agreed that there is no way to prove that one technique is better than any other. "While there is no agreement on a best method, ...some procedures are far more popular than others" (Jaeger, 1991, p. 491). Standard-setting procedures are based on pragmatics, not science. All of the techniques require that those setting the test standard impose their professional judgment to the task. To some (e.g., Glass, 1978), the judgments involved in these tasks are intrinsically arbitrary and therefore of questionable value. One reason that some testing professionals exhort caution in the standard setting process is that in choosing among the available standard setting techniques, one influences the standards to some degree. Similarly, the judges one uses in establishing test standards also impact the standards to a substantial degree (Jaeger, 1991).

The techniques employed in setting standards have been presented by Livingston and Ziecky (1982) and reviewed completely by Jaeger (1989); such detail is certainly beyond the scope of the present discussion. However, most graduation tests set standards by holding panels which review the test item-by-item to determine what the appropriate passing score should be. Such approaches are what Hambleton and Eignor (1980) call judgment models since they rely on the judgments of the panel members. The most common of these models is what has become called the Angoff procedure (after Angoff, 1971). In this procedure, a panel of judges is convened and each member of the panel reviews each question on the test and estimates the probability (a proportion from 0.00 to 1.00) that a minimally competent student would answer each correctly. These estimates are summed for each judge and then the individual judges' estimates are averaged. The resulting value becomes the passing score. The advantage of this procedure is that the passing score that is set is specific to the test in question and is based on judgments of those presumably knowledgeable to make such judgments. Among the disadvantages are the difficulties in determining what a "minimally competent" student would be, much less how he or she would perform on the test.

A few variations to the standard Angoff procedure may be employed. For example, one can have the judges themselves take the examination prior to their making judgments about the test questions. One can provide the judges with item analysis data so that they can see how test takers actually performed on each test question. One can also iterate the Angoff procedure several times with the same or different panels and provide each successive panel with the results of the preceding judgments. Another modification is to permit the judges to select a probability that a minimally competent test taker would answer a question correctly from a shortened list of the possible values from 0.00 to 1.00.

To be able to make ratings on specific items, as in an Angoff panel, a clear understanding of what minimum competence means is needed. Mills, Melican and Ahluwalia (1991) have addressed techniques to use with Angoff panelists to help them understand the multiplicity of different interpretations of the minimum competence concept. For example, the panel may begin by listing the levels of knowledge and skills that such an individual might possess. Those running the meeting need to keep the panelists focussed on the target individual -- a person graduating from high school with the least amount of knowledge and skills permissible. To be able to make such judgments, those making the ratings must be knowledgeable about the full range of skill levels of graduating seniors and about the curriculum and instruction that such students receive.

Jaeger (1991) has addressed the issue of who should compose a standard-setting panel. Standard 6.9 of the Standards for Educational and Psychological Testing (AERA et al., 1985) requires that the qualifications of judges composing the panel should be enumerated in an appropriate publication -- an implication that such factors have relevance. The task of serving on a standard-setting panel is a complex one involving the reading, understanding, and evaluating of a vast amount of detailed information typically in a confined time period. Judges must possess "substantial knowledge...that is rapidly accessible and readily integrated" (Jaeger, 1991, p. 3). Jaeger defines the ideal judges as experts.

> In the case of a high school graduation test, such individuals know the knowledge requirements of entry-level, post-high-school jobs or freshman courses in colleges and universities, assuming the purpose of high school graduation tests is to ensure that high school graduates possess knowledge sufficient to enter the labor force or enter a post-secondary education program. Judges most likely to possess this kind of expertise include directors of apprenticeship programs for craft unions, personnel directors of service-oriented companies that hire large numbers of recent high school graduates, college and university admission officers, and college and university faculty members who teach freshman courses (Jaeger, 1991, p. 4).

Mills et al., (1991) supplement Jaeger's recommendations by suggesting that "panelists in standard-setting studies should be chosen to represent all appropriate groups in the profession relevant to establishing the cutoff scores for the test. These panelists, therefore, will bring a diversity of knowledge, training, and opinions about the test and testing situation to the rating session" (Mills et al., p. 9). In the instance of setting standards that affect LEP students, such panels should probably include ESL instructors and others knowledgeable about the performance of LEP students.

It may be recalled that only one procedure for setting a standard has been described.

> A large number of empirical studies have addressed the question of whether different standard-setting procedures, when applied to the same competency test, provide similar results. Most research has answered this question negatively. Different standard-setting procedures generally produce markedly different test standards when applied to the same test, either by the same judges or by randomly parallel samples of judges (Jaeger, 1989, p. 497).

That different panels of judges and different procedures may elect to set varying standards has led some scholars (e.g., Jaeger, 1989; Shepard, 1980) to suggest using several methods in combination and then "consider all of the results, together with extrastatistical factors when determining a final cutoff score" (Jaeger, 1989, p. 497).

Adjustments made to initial standards. Geisinger (1991) has provided a list of some of the kinds of information that may be used to adjust the proposed passing scores that emerge from standard setting panel meetings. With respect to high school graduation tests, this information includes: (a) what passing rates/failing rates are acceptable to relevant parties; (b) the relative costs of misclassification errors (e.g., failing someone who should have passed); (c) societal needs; (d) adverse or disparate impact data; (e) errors of measurement due to the test's unreliability; (f) errors of rating due to differences among raters within a standard-setting panel and across different panels; (g) anomalies in the rating process (e.g., judges who are found to lack the expertise required of them); (h) how frequently and how often students are able to re-take forms of the examination; and (i) results of other standard-setting procedures. One can imagine several of these adjustments that are relevant for the assessment of LEP students. Most obvious, of course, is (d) adverse impact data. If the proportion of Hispanics passing the test, for example, is sufficiently below that of other groups, test makers, educational leaders and other concerned parties should review the results as well as the education of the students involved to consider what should be done. Perhaps some adjustment either to the overall passing point or the passing point for Hispanic test takers may be in order. A more subtle example concerns (e) test reliability. Passing scores are sometimes adjusted (typically in a downward direction) due to unreliability. Students who fall just below the passing score are seen as being strong contenders for passing the test, if it were only more reliable. The reliability coefficient and, more importantly, the standard error of measurement for LEP students taking the examination should be computed and compared to that of the majority students. If the reliability is lower and the standard error of measurement higher, an argument for a reduction in the passing score for LEP students would appear justifiable. As a final example, consider (c) societal needs. Paulson and Ball (1984) have argued that minorities were not as able to receive employment in the State of Florida after the high school graduation test was announced. Such information might argue that the test standard be reduced. On the other hand, if the results of testing are used to provide high quality remedial education to the LEP test takers who fail and this remedial education provides LEP students with improved academic skills without consequential personal, social, or academic costs (e.g., stigmas), then the competency test standard should be kept where it is or even increased.

There may be circumstances in the use of minimum competency examinations where it is appropriate to employ a different standard as the passing score than is used in the general population. In some instances, LEP students have already been identified for special test administration procedures such as being excluded from taking the examination altogether on the basis of an IEP or a similarly institutionalized policy, bypassing the first test administration for which they are eligible, having the test administered in their native or first language, or taking an alternative measure. Under such circumstances, it may also be appropriate to use a different passing score in the recognition that their more limited English skills inhibit their best performance. Padilla (1979) suggested a similar notion with regard to employment settings in noting that there are situations in which it is appropriate for job candidates to be essentially given "extra-credit" for being bilingual. "In job settings where such bilingualism is functionally related to job success, such credit is indeed appropriate, although it is rarely given in civil service settings, for example. Such bonuses, appropriately awarded because the language skills enhance job performance, should be clearly seen as additional to any other advantages provided to members of language minorities in the attempt to increase their representation in the work force, on campuses, in advanced instruction, etcetera. Credit for being bilingual (French/English) is appropriately provided to managers in the public service of Canada, for example" (Geisinger, in press, a).

Methodological Issues Specific to LEP Students

Test Bias

Test bias is intrinsically and closely tied to the concept of test validity because, like validity, it rests primarily upon inferences based on test scores. As in the case of all judgments of test validation, threats to validity threaten proper test score interpretation. Just as validation was dominated by the criterion-related approach until the last decade or two (see Geisinger, in press, b), so has the study of bias been dominated by the criterion-related approach. Many of the definitions of bias that have been traditionally provided are difficult to extract from the criterion-related validation paradigm.

One definition of bias (Cole & Moss, 1989) moves beyond the criterion-related model. This definition states:

> An inference is biased when it is not equally valid for different groups. Bias is present when a test score has meanings or implications for a relevant, definable subgroup of test takers that are different from the meanings or implications for the remainder of the test takers. Thus, *bias is differential validity of a given interpretation of a test score for any definable, relevant subgroup of test takers* (p. 205).

Such a definition, as is shown below in this paper, has implications for the competency testing of LEP students. With respect to the content validation approach, problems in making valid inferences may be based upon differences across groups with regard to the appropriateness of a given domain of content or testing format or for how well specific questions cover the content domain. For example, when Spanish-speaking tenth grade students write responses on an essay final examination in History, the quality of their responses may be limited by their ability to write the answer in English. A source of test score variance becomes English writing ability and inferences which assume that the scores are solely due to knowledge of History are incorrect. (This information has been cited in Geisinger, in press, a.)

Bias detection techniques. Test bias has been scientifically studied for several decades. Typically, reviews of test bias research subdivide the procedures which have been developed into external and internal methods. External methods are those that evaluate whether the relationship between test scores and extra-test criteria is comparable across groups. There are two types of internal methods. The first attempts to identify those test questions which are differentially more difficult for a given group than other questions composing the test. The second involves factor analyses of test items to identify dimensions of test performance for each of the groups under study. The attempt is to show that the test measures the same, similar or different characteristics across the varying groups. If similar factors are found across groups, we have some reason to suppose that the test measures comparable constructs in each group. This second approach is not discussed in the present paper but may be found in Geisinger (in press b), Reynolds (1982b), or Shepard (1982).

Reynolds (1982a) offered the following definition of test bias from the perspective of construct-related validation:

> Bias exists in regard to construct validity when a test is shown to measure different hypothetical traits (psychological constructs) for one group than another or to measure the same trait but with differing degrees of accuracy (p. 194).

Reynolds (1982b) suggested a number of different empirical techniques in which construct-related test bias might be identified. These include differences across reliability coefficients; rank ordering of item difficulties; correlations with other variables, such as age; comparisons of multitrait-multimethod matrices across groups; and factor-analytic differences. To know definitively whether competency tests measure the same cognitive processes for LEP and majority-group students requires research of this type. However, the numbers of LEP students, especially when subdivided by cultural or ethnic group, would likely prohibit such efforts.

In any test of cognitive skill or ability, a logical and elementary check that the test is comparable across groups is a determination that the relative difficulty of test questions is similar. That is, those questions which are difficult for one group should also be challenging for the other. Should such a finding not hold, one must question population validity of the underlying construct. If test items are rank ordered from easy to difficult within each group, rank order correlations, such as rho, may be calculated to demonstrate parity. Reynolds (1982b) suggests that rho's of .90 be taken as indicative of consistency of construct-related validity.

Test bias against LEP students. In some instances, the study of test bias against LEP students differs from that of other groups, such as African Americans, females, or the handicapped. The study of test bias is more difficult with language minorities because there are at least two ways in which such test bias differs from that against females, African Americans, and to a lesser extent, the handicapped. The first of these relates purely to language differences, both in test administration and in the interpretation of test results. The second situation considers differences between LEP students and the other groups in our society stemming from cultural factors, with these cultural differences including those related to language.

Two primary factors confound the interpretations of some tests with LEP students: language and culture. Neither of these problems is easy to deal with, but language has received more recent attention from the psychometric literature (e.g., Duran, 1988, 1989), although it is potentially less complex.

(1.) Language differences. The first issue related to language concerns the question, in which language should the test be administered (Geisinger, in press, a)? If the purpose of the test is to provide diagnostic information so that an IEP can be developed and instruction optimized, administering comparable competency tests in both languages may yield useful information.

When English-language tests are used with LEP students, the level of language applied on the test needs to parallel that used in the schools (as it does with English-speaking test takers). With written tests, various readability formulae may be used to estimate both the reading levels of the examination and of materials used in the classroom and in educational materials to ensure that the test does not require an artificially high reading level. Once again, the concept of differential instructional validity may be relevant if LEP students use educational materials which are generally easier to read than are the test materials. In the scoring of certain free-response measures, such as the essay examinations, the level of English language skill necessary for achieving passing scores on the examination should also be considered when constructing the questions, scoring the responses, and interpreting the results.

The 1985 Standards for Educational and Psychological Testing provide a section on the testing of linguistic minorities. Seven standards were enumerated to provide some guidance toward good testing practice. In general, these standards emphasize attempts to achieve valid inferences from test scores coming from members of linguistic minorities. They accent the notion that tests may be influenced by language skills (irrelevant to the construct purportedly being measured) to a greater or lesser extent when given to linguistic minorities. Thus, these standards attempt to assure valid test use and interpretation. Furthermore, they state that test constructors (in the case of minimum competency examinations), test publishers and state departments of education, developing assessment instruments recommended for use with linguistic minorities need to inform test administrators and test users of proper procedures and interpretations with those groups. The seven standards follow.

13.1 For non-native English speakers or for speakers of some dialects of English, tests should be designed to minimize threats to test reliability and validity that may arise from language differences.

13.2 Linguistic modifications recommended by test publishers should be described in detail in the test manual.

13.3 When a test is recommended for use with linguistically diverse test takers, test developers and publishers should provide the information necessary for appropriate test use and interpretation.

13.4 When a test is translated from one language or dialect to another, its reliability and validity for the uses intended in the linguistic groups to be tested should be established.

13.5 In employment, licensing, and certification testing, the English language proficiency level of the test should not exceed that appropriate to the relevant occupation or profession.

13.6 When it is intended that the two versions of dual-language tests be comparable, evidence of test comparability should be reported.

13.7 English language proficiency should not be determined solely with tests that demand only a single linguistic skill. (AERA et al., 1985, pp. 74-75).

The first six standards are clearly relevant to the testing of LEP students. Based on Roeber's (1990) survey, it would appear that few if any states are performing the research needed to ensure that different language test forms and other alternative test forms are comparable and equally valid. Without such information, equivalent interpretations may not be made using the different forms. Unfortunately, it may not be in the best interests of advocates of LEP students to demand such research. If states find it impossible to perform such research for budgetary, manpower, or other reasons, they may simply discontinue offering alternative testings or testings in a variety of languages.

(2.) Cultural differences. Gordon (1991) has defined culture as a complex whole that includes knowledge, belief, art, morals, law, custom and any other capabilities and habits acquired as a member of society. The total pattern of human behavior and its products -- embodied in thought, speech, action, and artifacts -- are dependent on the capacity to learn and transmit knowledge to succeeding generations through the use of tools, language, and systems of abstract thought. As a descriptive concept, culture is a product of human action; as an explanatory concept, it is seen as influencing further action (p. 101).

Scores emerging from tests are indeed subject to cultural influences. Not to reflect culture, of course, would likely mean that the scores do not validly reflect the construct or behavior they have been intended to assess. From two quotes by Anne Anastasi, we may glean when such influences represent valid influences upon test scores and when they do not. First, let us consider the valid perspective.

> Every psychological test measures a sample of behavior. Insofar as culture affects behavior, its influence will and should be reflected in the test. Moreover, if we were to rule out cultural differentials from a test, we might thereby lower its validity against the criterion we are trying to predict. The same cultural differentials that impair an individual's test performance are likely to handicap him (sic) in school work, job performance, or whatever other subsequent achievement we are trying to predict (Anastasi, 1967, p. 299).

Nevertheless, a "test may be invalidated by the presence of uncontrollable cultural factors. But this would occur only when the given cultural factor affects the test without affecting the criterion" (Anastasi, 1950, p. 15; also cited in Geisinger, in press, a).

Proper test score interpretation for linguistic minorities involves consideration of acculturation. "Acculturation refers to complex processes that take place when diverse cultural groups come into contact with one another. It is an extremely important aspect of the experience of linguistic minorities in the United States. Acculturation is also related to testing issues because it involves the acquisition of language, values, customs, and cognitive styles of the majority culture -- all factors that may substantially affect performance on tests" (Olmeda, 1981, p. 1082). Since acculturation can presently be assessed with substantial reliability and validity (Olmeda, 1979, 1981), teams planning the IEPs for LEP students should include formal measures of acculturation when making assessments of these students.

Item selection including item bias detection techniques. Item bias techniques are also known as methods to determine differential item functioning across groups, or DIF, and have been used in the pre-testing phase of test development to identify and remove those questions from a test that are differentially more difficult for one group or another. The manner in which most of the available techniques work may be explained as follows. The only two factors employed by these techniques are the group-specific difficulty level of each of the questions composing the test and the overall level of ability or knowledge of each test taker. The test taker's level of ability or knowledge is generally designated by the individual's overall score on the examination or some mathematical derivation of this value.

The logic of the process is that both the content and thought processes called for by a question determine the difficulty level of that question and should be comparable across groups. Since groups may differ in terms of overall ability for one reason or another, these techniques adjust difficulty levels of individual test questions for these overall group differences. If the difference between difficulty levels for an item for two groups is disproportionally large, even given the groups' overall test differences, the particular test question is considered biased. In the pre-testing of an examination, such questions would likely be removed from the instrument under development.

Issues Relating to Standard Setting

It was reported previously in this paper that the most common technique for setting standards is to bring together a panel which reviews the test on a question-by-question basis and generates data used to set the passing standard. Representation by LEP parents, educators or other delegates should appear on these panels. This representation would permit discussion among the members of the panel of issues of relevance to LEP students. It should be openly questioned as to whether a single standard is appropriate or rather if differential standards should be applied for varying groups of students. In cases where a single standard setting was held at some previous time and the state now simply equates the passing score of new tests to this pre-existing standard, either new standard setting panels should be convened or adjustments to the standard should be considered.

Equating

Scores on one form of a state's minimum competency test are frequently equated to previously used forms so that test scores and passing scores retain their meaning over time. Equating generally involves a sample of students taking part in all of both forms of the examination. It is critical that LEP students be included in representative numbers in these equating samples. In fact, given their small numbers in many locations, they may need to be greatly oversampled.

Furthermore, it is unlikely that states would be able to equate special language versions of the test if they are available. Equating methodology would generally require either that randomly drawn, equivalent groups of individuals take both versions of the examination or that if the same group took both test forms, their language skills be equal in both languages. These assumptions will almost never be met. In addition, the score distributions emerging from English-language tests and foreign language tests are unlikely to parallel each other, with the distribution of foreign test scores well below that of the English-speaking test takers.

Instructional Feedback

Competency testing laws in many states require that those students who fail the minimum competency test receive remedial instruction so that they may succeed when they re-take the examination. Ultimately, the key to the successful competency testing of LEP students involves a proper diagnosis of their academic weaknesses and strengths as well as the development of a well formulated educational plan to remediate their shortcomings in an optimal manner. A successful diagnostic test must be sensitive to their instruction at a micro-level (see Duran, in press) and able to yield reliable information about exactly what the students can and cannot do. Although minimum competency tests are generally group tests, they need to be interpreted on an individual basis, especially for LEP students. That is, educational professionals need to consider the test data along with other indices of educational performance (e.g., work in class), academic skills (e.g., strengths and weaknesses in English, their native language, and other academic information), and knowledge of the setting, broadly defined, in which the child may be found. Furthermore, remedial programs are likely to be negatively tainted and to have adverse impact against ethnic and language minorities. The remedial program in which LEP students are placed must be successful not only educationally but also in terms of overcoming such a stigma. To the extent that the tests and their use are well integrated into the instructional program, they may prove to be successful.

Advantages and Disadvantages of Minimum Competency Testing

In that minimum competency programs have been in effect in one manner or another for more than 20 years, it is disappointing that no comprehensive evaluation studies of these programs have appeared in the literature. If they exist in the files of states, they need desperately to be shared. With the lack of formal evaluations, we must hypothesize and reflect upon the potential advantages and disadvantages of these testing programs from the armchair. Formal evaluations of these testing programs would be strongly recommended before the federal government moves to operationalize the idea of national examinations.

Societal Effects

One advantage to society if the ideal of minimum competency testing were realized would be that society would become filled with adults each of whom is able to read, write, and use basic mathematical skills. Based on the requirements of some states, graduates would also be able to communicate orally (speak, listen, etcetera) effectively.

From a more negative perspective, the success of minority students on minimum competency tests, as on other examinations, is far below that desired (e.g., Hartigan & Wigdor, 1989). "In states such as North Carolina that maintain statistics on the characteristics of students who fail competency tests, the failure rates of racial minorities are typically found to be 5 to 10 times higher than those of the majority white students. The social and economic consequences of failing to earn a high school diploma are well-known, particularly for youths from minority groups (cf., Eckland, 1980)" (Jaeger, 1989, p. 491). In light of these performance differences, test bias procedures, as mentioned earlier in this paper, may need to be applied. Industrial testing, as opposed to educational testing, has been forced to study the impact of testing when rewards are assigned on the basis of test scores.

The Uniform Guidelines on Employee Selection (1978), issued jointly by the Equal Employment Opportunity Commission, the then Civil Service Commission, the Department of Labor, and the Department of Justice, after considerable input from professional organizations interested in testing practice, operationalized good testing practice in industrial settings in many ways. The Guidelines defined a model of proper test use in which a test need only be shown to be valid for the use to which it is being put after it has first been shown to have adverse impact upon a protected group (defined as Blacks; American Indians; Asians including Pacific Islanders; Hispanics including persons of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish origin or culture; women and other groups). Adverse impact has generally been defined by the "four-fifths rule." That is, "a selection rate for any race, sex, or ethnic group which is less that four-fifths (4/5) (or 80 percent) of the rate for the group with the highest rate will generally be regarded...as evidence of adverse impact" (3D).

The use of minimum competency tests has the potential for real test misuse. Consider the following comments that were distributed to users of the Graduate Record Examination (Graduate Record Examination Board, 1989) as the first of 10 recommended guidelines for proper test use. The guideline states:

> Regardless of the decision to be made, multiple sources of information should be used to ensure fairness and balance the limitations of any single measure of knowledge, skills, or abilities.... Scores should not be used in isolation. Use of multiple criteria is particularly important when using...scores to assess the abilities of educationally disadvantaged students, students whose primary language is not English, and students returning to school after

> an extended absence. Score users are urged to become familiar with factors affecting score interpretation for these groups (Graduate Record Examination Board, 1989, p. 6).

Four of the other nine guidelines are also relevant to competency testing. The four specific guidelines suggest that validity studies need be performed, test content be reviewed by subject matter experts, decisions based on small score differences be avoided and test users recognize limitations of scores earned on tests taken under special administrative arrangements (e.g., in a language other than English).

Effects on Students

Blau (1980), a clinical psychologist, has considered what the psychological effects of institutionalized minimum competency testing programs are likely to be. It should be noted from the outset that his sample was small (around 35 students) and apparently gathered from the many adolescents that he saw in his clinical practice. In some cases he was seeing them specifically because of their educational difficulties. Relevant to the present discussion, he also does not report if any of the members of his sample are linguistic minorities. Nevertheless, he reports that the students were "distressed and disdainful about the whole testing business. They saw it as another burden developed by adults to make their progress through school more difficult" (p. 176). With regard to their performance on the test, in this case the Florida high school graduation test, he reported that "the majority of the students, including the very bright ones, simply do not care" (p. 176). The rationale for their apathy was described as differing depending upon how strong the students were. "The poor students saw the tests as an additional barrier to success and esteem and not a help, while the good students saw them as a barrier to using time effectively" (p. 177). One factor appeared to moderate the involvement of students: the immediacy of the feedback that they received. When such feedback was received quickly by students, they did see it as of educational value. In attempting to address how such overly negative attitudes toward the competency testing process might be addressed, Blau called for the involvement of (representatives of) students involved in every stage of the testing process.

One problem that may beset students relates to their moving from one school district or one state to another. Suddenly, different requirements or higher standards impact students. Such problems would be especially notable in situations where school districts set district-level standards for passing statewide examinations. A student might move only a few blocks but, on that account, fail to pass an examination that he or she had apparently already cleared.

Effects on Schools and the Educational Process

The effects of minimum competency testing on the educational process can be considered from a variety of perspectives. One relates to the ultimate goal of the examinations in general. According to Jaeger (1989, pp. 486-87), "Although some competency testing programs attempt to inform students about their academic strengths and weaknesses, the principal use of competency-test results is to serve institutional purposes such as student placement rather than individual purposes such as student guidance and counseling."

One of the biggest fears of those reluctant to endorse minimum competency testing concerns the notion that the low level or minimal skills frequently measured on these tests will become the maximal skills taught by educators. That is, it is feared that teachers will stop striving to teach higher-level thinking and problem-solving skills as long as their students master those basic, life skills called for by the minimum competency examinations.

Legal Issues

Reviews of court records have indicated that courts have continually upheld the rights of states to employ minimum competency tests to monitor the success of educational programs and the skill levels of potential high school graduates (Citron, 1982, 1983; Jaeger, 1989).

The most influential case regarding minimum competency testing brought to date was *Debra P. v. Turlington* (1979, 1981, 1983, 1984). This federal case received considerable attention as evidenced by George Madaus' (1982) book dedicated to the history, effects, and implications of the case. The case, which related to Florida's high school graduation test, was brought by 10 African-American students who had failed the examination and who challenged the adverse impact of the examination against the backdrop of a long history of segregated schools and other forms of discrimination. Florida's minimum competency examination was a test of functional literacy which had been mandated by a 1976 statute requiring demonstration of such skills in order to receive a high school diploma. Functional literacy was defined as skills in reading, writing, and arithmetic needed to face successfully problems encountered in everyday adult life. Reading and writing were combined as a test of communication skills. In its first administration in 1977, 36 percent of high school seniors failed one or both of the examinations, but 77 percent of African-American students failed against 24 percent of white students (Pullin, 1982). "After three attempts to pass the test, 1.9 percent of white students and 20 percent of black students...still failed" (Jaeger, 1989, p. 507) the test. Two sets of claims were made against the test. First, it was argued that the test was discriminatory on the basis of its adverse impact; it breached the constitutional rights of equal protection under the laws as guaranteed by the Fourteenth Amendment and as enforced under Title VI of the Civil Rights Act of 1964, which prohibited discrimination on the basis of race, among other factors. Since disparate impact had been found in this instance, Florida needed to prove that the differences in passing rates had not been caused by the state's history of discrimination. Second, it was contended that the test was invalid and the requirement that it be passed to earn a high school diploma hastily conceived.

The trial iterated several times between the federal district court level and the appellate court over a period of five years. Initially, a time-period delimited moratorium was placed on the use of the test so that students in the schools were able to prepare for it more properly. The principles of curricular and instructional validity advanced by McClung (1978, 1979) were critical in that the state of Florida was ordered to document that students in Florida schools universally received instruction in the content represented by the tests. The courts initially ruled that punishing the victims of past discrimination "for deficits created by an inferior educational environment neither constitutes a remedy nor creates better educational opportunities" (474 F. Supp., at 257, *Debra P. v. Turlington*, 1979; also cited in Jaeger, 1989). They ultimately ruled, however, based on an overwhelming amount of data indicating that the content of the tests was covered both in curricula throughout the state as well as in actual instruction, that the tests were both fair and valid.

In summary, courts have upheld the rights of states to use competency tests appropriately but have placed limitations on the testing programs (1) when there is a history of discrimination, (2) when students have not been given adequate advance warning about the necessity to pass the tests, and (3) when the curriculum and the instruction provided do not cover the material on the test (Jaeger, 1989).

Using Minimum Competency Tests with LEP Students

Competency tests, like other educational tests, have the capacity to improve the education of the students in our country's schools. To be effective, however, they need to be linked closely to instruction. That is, they need to have high instructional and curricular validity. Furthermore, the curriculum needs to drive the content of the examinations rather than vice versa. One of the most damning indictments of all educational tests is that they determine what is taught in some instances. It may be noted that teaching to a test is not always bad. Providing high quality instruction on topics of high relevance and importance, of course, will always be paramount. However, decisions as to what should be taught are curriculum-level decisions that should be made when developing the curriculum and instructional approach, not after determining what is on an examination.

Making minimum competency tests instructionally meaningful involves more than curricular and instructional validity of the examinations, however. It also entails using the scores to provide access to remedial instruction rather than "as a stick" to punish those who fail, perhaps by withholding diplomas or other valued rewards (see Serow, 1984).

It has been stated that about 2 percent of the students who take minimum competency tests do not pass them, even after repeated administrations of forms of the examinations. It has been further argued by others (Serow, 1984) that this small percentage is politically acceptable to the policy makers who in some cases recommend or require the examinations. Serow reminds us that small percentages of large bodies are still a large number of failures. One wonders if these small numbers would be acceptable if they were being added to welfare rolls rather than being refused a high school diploma.

One issue in discussing the competency testing of LEP students is not a testing matter at all, but purely an educational one. Perhaps all LEP students should have IEPs just as other exceptional students and students with handicaps already do. LEP students are among the most disadvantaged students not so covered at the present time. Should they be provided with the planning and supportive remediation required by an IEP? With such attention, the success rate of LEP students would surely rise.

Summary

Deciding between withholding a diploma from a student who has spent 12 years in ineffective schooling and graduating a student who lacks basic academic and life skills is a no-win choice. The only acceptable solution to this decision is to use the test scores to identify students needing remedial instruction. The most useful such test would be one that is diagnostic rather than summative. However, most statewide competency tests are by their very nature summative tests that do not provide diagnostic information.

One must question whether a minimum competency test can possibly be equally valid from the perspective of curricular and instructional validity and not biased for LEP students, on the very basis of their differential needs and educational programs.

For competency tests to be most useful for improving the education of LEP students, it is imperative that the tests be closely tied to the curriculum, be thoroughly integrated with the curriculum, aim toward providing diagnostic instructional and remedial feedback, provide scores which are readily interpretable by educational professionals, and become less threatening than they appear to have become. Failing scores on competency examinations need to be attuned to the development of IEPs for those LEP students requiring them. The notion that all LEPs would benefit from IEPs has some merit and should be investigated.

The psychometric literature coupled with pragmatic realities of the situation have little to offer at the present time with regard to ways of determining (1) whether minimum competency tests are as valid for LEP students as for others and (2) what passing scores should be used for such students.

Interpretation of individual test scores is extremely demanding. Complex interactions of psychological, language, culture, and other background factors affect the test performance of linguistic minorities. Examiners and educational planners need to be specially trained to test such individuals and to consider language skills, acculturation, socioeconomic factors and other variables in any assessment of an individual's level of functioning.

Notes

[1] The author would like to thank Scott Cone for his help on this paper, Janet F. Carlson for her careful reading of the paper, and Michael Beck of Beck Evaluation Testing Associates, Chris Pipho of the Education Commission of the States and Ed Masonis of the New Jersey Department of Education for their helpful information. Any errors in this paper, of course, remain those of author.

[2] The information provided in this paragraph was taken from Roeber's (1990, pp. 17-18) survey of statewide testing practices.

References

American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Anastasi, A. (1967). Psychology, psychologists, and psychological testing. American Psychologist, 22, 297-306.

Anastasi, A. (1950). Some implications of cultural factors for test construction. In Proceedings of the 1949 ETS Invitational Conference (pp. 13-17). Princeton, NJ: Educational Testing Service.

Anastasi, A. (1988). *Psychological testing* (6th ed.) New York: Macmillan.

Anastasi, A. (1990). What is test misuse: Perspectives of a measurement expert. In *The uses of standardized tests in American education: Proceedings of the 1989 ETS Invitational Conference* (pp. 15-25). Princeton, NJ: Educational Testing Service.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.; pp. 508-600). Washington, DC: American Council on Education.

Berk R. A. (1984). Selecting the index of reliability. In R. A. Berk (Ed.), *A guide to criterion-reference test construction* (pp. 231-266). Baltimore: John Hopkins University Press.

Blau, T. H. (1980). Minimum competency testing: Psychological implications for students. In R. M. Jaeger & C. K. Tittle (Eds.), *Minimum competency achievement testing: Motives, models, measures, and consequences* (pp. 172-181). Berkeley, CA: McCutchan.

Brennan, R. L. (1984). Estimating the dependability of the scores. In R. A. Berk (Ed.), *A guide to criterion-reference test construction* (pp. 292-334). Baltimore: John Hopkins University Press.

Citron, C. H. (1982). Competency testing: Emerging principles. *Educational measurement: Issues and practice*, *1*, 10-11.

Citron, C. H. (1983). Courts provide insight on content validity requirements. *Educational measurement: Issues and practice*, *2*, 6-7.

Cloud, N. (1991). Educational assessment. In E. V. Hamayan & J. S. Damico (Eds.), *Limiting the bias in the assessment of bilingual students* (pp. 219-246). Austin, TX: Pro-ed.

Cohen, D. K. & Haney, W. (1980). Minimum competency testing and social policy. In R. M. Jaeger & C. K. Tittle (Eds.), *Minimum competency achievement testing: Motives, models, measures, and consequences* (pp. 15-22). Berkeley, CA: McCutchan.

Cole, N. S. & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 201-220). New York: American Council on Education & Macmillan.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.) (pp. 443-507). Washington, DC: American Council on Education.

Debra P. v. Turlington, 474 F. Supp. 244 (M.D. Fla. 1979).

Debra P. v. Turlington, 653 F. 2nd 397 (5th Cir. 1981).

Debra P. v. Turlington, Case 78-792, Memorandum Opinion and Order (M.D. Fla., May 4, 1983).

Debra P. v. Turlington, Case 83-3326 (11th Cir. C. App., April 27, 1984).

Donlon, T. F. (In press). Legal issues in the testing of Hispanics. In K. F. Geisinger (Ed.), *The psychological testing of Hispanics*. Washington, DC: American Psychological Association.

Duran, R. P. (1983). *Hispanic's education and background: Predictors of college achievement*. New York: College Entrance Examination Board.

Duran, R. P. (1988). Validity and language skills assessment: Non-English background students. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 105-128). Hillsdale, NJ: Lawrence Erlbaum Associates.

Duran, R. P. (1989). Testing of linguistic minorities. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 573-588). New York: American Council on Education & Macmillan.

Duran, R. P. (In press). Clinical assessment of instructional performance in cooperative learning. In K. F. Geisinger (Ed.), *The psychological testing of Hispanics*. Washington, DC: American Psychological Association.

Eckland, B. K. (1980). Sociodemographic implications of minimum competency testing. In R. M. Jaeger & C. K. Tittle (Eds.), *Minimum competency achievement testing: Motives, models, measures, and consequences* (pp. 124-135). Berkeley, CA: McCutchan.

Finch , F. L. (Ed.) (1991). *Educational performance assessment*. Chicago: Riverside.

Gallagher, J. J. (1980). Setting educational standards for minimum competency: A case study. In R. M. Jaeger & C. K. Tittle (Eds.), *Minimum competency achievement testing: Motives, models, measures, and consequences* (pp. 239-257). Berkeley, CA: McCutchan.

Geisinger, K. F. (1991). Using standard-setting data to establish cutoff scores. *Educational Measurement: Issues and Practice*, *10* (2), 17-22.

Geisinger, K. F. (In press, a). Fairness and selected psychological issues in the testing of Hispanics. In K. F. Geisinger (Ed.), The psychological testing of Hispanics. Washington, DC: American Psychological Association.

Geisinger, K. F. (In press, 1992). The metamorphosis in test validation. Educational Psychologist. Also published in The Educational Researcher, 1991, 29(1), 2-12.

Glass, G. V (1978). Standards and criteria. Journal of Educational Measurement, 15, 237-261.

Gordon, E. W. (1991). Human diversity and pluralism. Educational Psychologist, 26, 99-108.

Gorth, W. P. & Perkins, M. R. (1979). A study of minimum competency testing programs. Amherst, MA: National Evaluation Systems.

Hamayan, E. V. & Damico, J. S. (1991). Limiting the bias in the assessment of bilingual students. Austin, TX: Pro-ed.

Hambleton, R. K & Eignor, D. R. (1980). Competency test development, validation, and standard setting. In R. M. Jaeger & C. K. Tittle (Eds.), Minimum competency achievement testing: Motives, models, measures, and consequences (pp. 367-396). Berkeley, CA: McCutchan.

Haney, W. (1982). Validity and competency tests: The Debra P. Case, Conceptions of validity, and strategies for the future. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing (p. 63-94). Boston: Kluwer-Nijhoff Publishing.

Hardy, R. A. (1984). Measuring instructional validity: A report of an instructional validity study for the Alabama High School Graduation Examination. Journal of Educational Measurement, 21, 291-304.

Hartigan, J. A. & Wigdor, A. K. (Eds). (1989). Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery. Washington, DC: National Academy Press.

Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 485-514). New York: American Council on Education & Macmillan.

Jaeger, R. M. (1991). Selection of judges for standard-setting. Educational Measurement: Issues and Practice, 10 (2), 3-6, 10, 14.

Jaeger, R. M., & Title, C. K. (Eds.). (1980). Minimum competency achievement testing: Motives, models, measures, and consequences. Berkeley: McCutchan.

Linn, R. L. (1982). Curricular validity: Convincing the court that it was taught without precluding the possibility of measuring it. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing. (pp. 115-132). Boston: Kluwer-Nijhoff Publishing.

Linn, R. L., Madaus, G. F., & Pedulla, J. J. (1982). Minimum competency testing: Cautions on the state of the art. American Journal of Education, 91, 1-35.

Livingston, S. A. & Ziecky, M. J. (1982). Passing scores: A manual for setting standards of performance on educational and occupational tests. Princeton, NJ: Educational Testing Service.

Madaus, G. F. (1982a). Competency testing: State and local level responsibilities. Educational Measurement: Issues and practice, 1, 10.

Madaus, G. F. (Ed.) (1982b). The courts, validity, and minimum competency testing. Boston: Kluwer-Nijhoff Publishing.

Madaus, G. F. (1982c). Minimum competency testing for certification: The evolution and evaluation of test validity. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing (pp. 21-62). Boston: Kluwer-Nijhoff Publishing.

McClung, M. S. (1978). Are competency testing programs fair? Legal? Phi Delta Kappan, 57, 397-400.

McClung, M. S. (1979). Competency testing programs: Legal and educational issues. Fordham Law Review, 47, 651-712.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 30, 955-966.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35, 1012-1027.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 13-103). New York: American Council on Education & Macmillan.

Mills, C. N., Melican, G. J., & Ahluwalia, N. T. (1991). Defining minimum competence. Educational Measurement: Issues and Practice, 10(2), 7-10.

Oller, J. W., Jr. & Damico, J. S. (1991). Theoretical considerations in the assessment of LEP students. In E. V. Hamayan & J. S. Damico (Eds.), Limiting the bias in the assessment of bilingual students (pp. 77-110). Austin, TX: Pro-ed

Olmeda, E. L. (1979). Acculturation: A psychometric perspective. American Psychologist, 34, 1061-1070.

Olmeda, E. L. (1981). Testing linguistic minorities. American Psychologist, 36, 1078-1085.

Padilla, A. M. (1979). Critical factors in the testing of Hispanic Americans: A review and some suggestions for the future. In R. W. Tyler & S. H. White (Eds.), Testing, teaching and learning: Report of a conference on research on testing (pp. 219-233). Washington, DC: National Institute of Education.

Paulson, D., & Ball, D. (1984). Back to basics: Minimum competency testing and its impact on minorities. Urban Education, 19, 5-15.

Reynolds, C. R. (1982a). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology. New York: Wiley.

Reynolds, C. R. (1982b). Methods for detecting construct and predictive bias. In R. A. Berk (Ed.), Handbook of methods for detecting bias (pp. 199-227). Baltimore, MD: Johns Hopkins University Press.

Roeber, E. D. (1990). Survey of large-scale assessment program: Fall, 1990. Lansing, MI: Michigan Educational Assessment Program, Michigan Department of Education.

Schmidt, W. H., Porter, A. C., Schwille, J. R., Floden, R. E., & Freeman, D. J. (1982). Validity as a variable: Can the same certification test be valid for all students? In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing. (pp. 133-150). Boston: Kluwer-Nijhoff Publishing.

Serow, R. C. (1984). Effects of minimum competency testing for minority students: A review of expectations and outcomes. The Urban Review, 16, 67-75.

Shepard, L. A. (1982). Definitions of bias. In R. A. Berk (Ed.), Handbook of methods for detecting bias (pp. 9-30). Baltimore, MD: Johns Hopkins University Press.

Shepard, L. A. (1980). Technical issues in minimum competency testing. Review of Research in Education, 8, 30-82.

Subkoviak, M. J. (1984). Estimating the reliability of mastery-nonmastery classifications. In R. A. Berk (Ed.), A guide to criterion-reference test construction (pp. 267-291). Baltimore: John Hopkins University Press.

Tittle, C. K. (1982). Competency testing. In H. E. Mitzel (Ed.), Encyclopedia of educational research (fifth edition) (pp. 333-352.). New York: Macmillan and the Free Press.

Trent, C. C., & Hinton, M. W. (1984). Mainstreaming, minimum competency testing and the minority student. The Western Journal of Black Studies, 8, 167-171.

Uniform guidelines on employee selection. (1978). Federal Register, 43(166), 38296-38309.

U. S. Department of Education. (1991). America 2000: An educational strategy. Washington, DC: Author.

Walker, D. F. What constitutes curricular validity in a high-school-leaving examination? In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing. (pp. 169-180). Boston: Kluwer-Nijhoff Publishing.

Willig, A. C. & Ortiz, A. A. (1991). The nonbiased individualized educational program: Linking assessment to instruction. In E. V. Hamayan & J. S. Damico (Eds.), Limiting the bias in the assessment of bilingual students (pp. 281-303). Austin, TX: Pro-ed.

Other Relevant Readings

Anastasi, A. (1958). Differential Psychology (3rd ed.). New York: Macmillan.

Baecher, R. E. (1982). The instruction of Hispanic American Students: Exploring their educational cognitive styles. In J. A. Fishman & G. D. Keller (Eds.), Bilingual education for Hispanic students (pp. 368-383). New York: Teachers College Press.

Barton, J. I. & Robinette, C. L. (1980). Minimum competency testing of pupils: Psychological implications for teachers. In R. M. Jaeger & C. K. Tittle (Eds.), Minimum competency achievement testing: Motives, models, measures, and consequences (pp. 155-171). Berkeley, CA: McCutchan.

Bernstein, D. K. (1989). Assessing children with limited English proficiency: Current perspectives, Topics in Language Disorders, 9(3), 15-20.

Broudy, H. S. (1980). Impact of minimum competency testing on curriculum. In R. M. Jaeger & C. K. Tittle (Eds.), Minimum competency achievement testing: Motives, models, measures, and consequences (pp. 108-117). Berkeley, CA: McCutchan.

Calfee, R. (1982). Establishing instructional validity for minimum competency programs. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing (pp. 95-114). Boston: Kluwer-Nijhoff Publishing.

Chamberlain, P., & Medeiros-Landurand, P. (1991). Practical considerations for the assessment of LEP students with special needs. In E. V. Hamayan & J. S. Damico (Eds.), Limiting the bias in the assessment of bilingual students (pp. 111-156). Austin, TX: Pro-ed.

Easton, L. B. (1991). Developing educational performance tests for a statewide program. In F. L. Finch (Ed.), Educational performance assessment. Chicago: Riverside.

Feldmesser, R. A. (1980). Minimum competency as an individual right. In R. M. Jaeger & C. K. Tittle (Eds.), Minimum competency achievement testing: Motives, models, measures, and consequences (pp. 425-437). Berkeley, CA: McCutchan.

First, J. M. & Cardenas, J. (1986). A minority view on testing. Educational Measurement: Issues and Practice. 5(1), 6-11.

Flaugher, R. L. (1978). The many definitions of test bias. American Psychologist, 33, 671-679.

Hunter, J. E. Schmidt, F. L. & Rauschenberger, J. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C. R. Reynolds, & R. T. Brown (Eds.), Perspectives on bias in mental testing (pp. 41-100). New York: Plenum.

Jensen, A. R. (1980). Bias in mental testing. New York: Free Press.

Katzenmeyer, W. G. & Steener, A. J. (1977). Estimation of the invariance of factor structures across race and sex with implications for hypothesis testing. Educational and Psychological Measurement, 37, 111-119.

Kretschmer, R. E. (1991). Exceptionality and the Limited English Proficient student: Historical and practical contexts. In E. V. Hamayan & J. S. Damico (Eds.), Limiting the bias in the assessment of bilingual students (pp. 1-38). Austin, TX: Pro-ed.

Leinhardt, G. (1982). Overlap: Testing what is taught. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing (pp. 151-169). Boston: Kluwer-Nijhoff Publishing.

Linn, R. L. (1982). Two weak spots in the practice of criterion-referenced measurement. Educational measurement: Issues and practice, 1, 12-13,25.

Nuttall, E. V. (1987). Survey of current practices in the psychological assessment of Limited-English-Proficiency Handicapped Children, Journal of School Psychology, 25, 53-61.

Pennock-Roman, M. (1990). Test validity and language background: A study of Hispanic American Students at Six Universities. New York: College Entrance Examination Board.

Reynolds, C. R., & Brown, R. T. (1984). Bias in mental testing: An introduction to the issues. In C. R. Reynolds & R. T. Brown (Eds.), Perspectives on bias in mental testing (pp. 1-40). New York: Plenum.

Samuda, R. J. (1975). Psychological testing of American minorities: Issues and consequences. New York: Dodd, Mead & Company.

Schmidt, F. L., Pearlman, K., & Hunter, J. E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. Personnel Psychology, 32, 257-281.

Steward, A. A., & Bernazza, H. (1982). Cloze procedure with Spanish, English and Bilingual adults. In J. A. Fishman & G. D. Keller (Eds.), Bilingual education for Hispanic students (pp. 326-332). New York: Teachers College Press.

Thorndike, R. L. (1971). Concepts of culture fairness. Journal of Educational Measurement, 8, 63-70.

Response to Kurt Geisinger's Presentation

Michele R. Hewlett-Gomez
Sam Houston State University, Texas

Response: Last July, Former Labor Secretary William Brock stated in a Time Magazine interview, "We are the only country in the industrial world that says to one out of every four of its young people, "We are going to let you drop out of sight. We are not going to give you the tools to be productive" [p.12, 1990].

Competency testing has been one answer for school districts in preparing students for the real world to become productive citizens. Competency testing began with the back to basic movement with the intentions to assess minimum skills as deemed necessary by some for a student to function in the real world. Does mastery of identified competency skills guarantee a society that high school graduates will possess the necessary skills to be productive citizens in the real world? This question ponders policy makers, educators, and parents in our efforts to determine a meaningful education for students and especially for limited English proficient (LEP) students within our public schools.

Today's topic thus drives the question of relevancy on evaluating students' learning outcomes and on a student's individual merit of success. Dr. Kurt Geisinger, in his paper, clearly presents the current testing issues facing policy makers, educators, parents, and test publishers in our public schools and, in particular, the issue of linking competency testing to a high school diploma for limited English proficient students. Dr. Kurt Geisinger brings to the forefront the urgency to reevaluate the purpose of minimum competency testing for students and for limited English proficient students, in light of our public education goals.

Dr. Kurt Geisinger precisely identified seven subtopics related to minimum competency tests and limited English proficient students. These thoroughly researched subtopics included minimum competency and its current status, methodological issues of the tests, methodological issues of the tests and LEPs, testing standards, advantages/disadvantages, legal issues, usage of tests with LEPs, and finally solutions. In essence of time, five subtopics are highlighted with the intention to arouse and stimulate further discussion.

1. **Current Status of Minimum Competency Testing** Dr. Kurt Geisinger used Roeder's national survey (1990) on the current status of minimum competency testing to survey states testing standards and found that the majority of states use a version of

competency testing to assess academic skills. Approximately 37 of the 50 states, responding to Roeber's survey, stated either criterion referenced, norm-referenced or performance tests were administered in a variety of disciplines ranging from reading, writing, and mathematics to citizenship, health, and fine arts. Students were generally tested at alternating grades every two to three years beginning at the third grade. In addition, Valdez-Pierce (1991) surveyed the southern states to find Mississippi and North Carolina administered minimum competency tests, totalling 39 states.

With the state's functions on competency testing varying due to diversities of populations, policies and educational philosophies, states tended to place their priorities on functions such as general information [34], followed by system accountability [28], and then curriculum mastery [20]. Interestingly, these functions lead one to ask, "why would a state demand schools to test students for general information?," "how does accountability correlate to a high school diploma?," "what are states teaching students when the testing program is not aligned to curriculum?," or "how many limited English proficient students do pass the minimum competency test?"

One function, system accountability, seemed to be a way for states to guarantee to society that districts will be accountable for student learning outcomes. Accountability issues generally stem from state governments that use the data on student performance to reward, punish, and/or assist schools. Obviously, these states take accountability seriously and attempt to improve student performance on the indicators stressed by the state government. The effect has been in many schools to deemphasize instructional quality, narrow the curriculum, and emphasize mastery of the test objectives.

For example, since 1981, Texas legislation has mandated testing minimum skills in mathematics, reading, and writing; first with the Texas Assessment of Basic Skills and then with the Texas Educational Assessment of Minimal Skills. In October 1990, a new criterion-referenced testing program, Texas Assessment of Academic Skills [TAAS], was implemented to provide a shift from an assessment of minimum skills to an assessment of academic skills with the intention to assess higher level thinking skills and problem solving abilities. What the Texas state government had discovered about its testing program over the past ten years was that students had not achieved sufficient mastery of objectives in mathematics, reading, or writing that address higher level thinking skills and problem solving abilities. For example, in the Grade 11 writing composition test, which uses analytic and holistic scoring, students were able to compose and sequence thoughts, yet lacked the ability to support and/or elaborate their thoughts in a composition. In the Grade 11 reading test, students scored below 70 percent mastery on objectives relating to summarization and points of view. In the Grade 11 mathematics test, students' mastery of some conceptual development, operational, and problem-solving skills were also consistently below 70 percent and in some cases below 60 percent such as in problem-solving using solution strategies. In effect, the testing program had shifted schools' standards for student success and accountability by rewarding districts for student success and punishing student with failures, alias no high school diploma. The ownership for student failure then had been on the student rather than the school system.

A second function, curriculum mastery, was used by states as a vehicle to align curriculum to an instructional and testing program. The relevancy of students' mastering specific skills became important, especially in the efforts to equal differential learning opportunities. Illinois and Minnesota were two states that developed their competency tests comprised of subject matter teams of teachers and curriculum experts. Texas, also, mandates a core-curriculum from which the districts design instructional programs to pass the state testing program. In addition, Roeber's survey identified Alberta, Canada, as using a minimum competency test, developed by teacher committees and scored by teachers, as partial criteria to receive a high school diploma. The test results in all three cases differ in their usage for program evaluation and instructional improvement. Should teachers be able to design their own testing programs to align with the curriculum and instructional program that addresses their student population? The issue of a decentralized authority to grant teachers control of curriculum for their student's learning merits priority consideration.

National. Besides the test functions, Dr. Kurt Geisinger addressed the current status of competency testing at the national level. The issue of a national test, American Achievement Test, is being called for by the Department of Education in the AMERICA 2000 report [U.S. Department of Education, 1991]. Since the purposes of this test are not clear, it then becomes even more critical that experts in the field of first and second language learning and testing present the relevant issues to policy making committees to define the uses of such unrealistic and misguided testing program. On April 23, 1991, Edward De Avila, an expert in linguistics and psychometrics, addressed the House Subcommittee on Select Education to state "That the development of a national test today was clearly to put the cart before the horse" (CTB News, 1991). His rationale was based upon the unfairness of administering a test to groups of children who had not received the same instruction, which would dictate local curriculum and, in particular, which would not necessarily tell us anything new about how students perform. Yet more importantly, De Avila focuses on the problem of definition. The lack of a consistent definition for limited English proficiency, compounds the tendency to pile all LEP students in one group regardless of the varying language and academic skills. De Avila states "that without a clear understanding [definition], we have no way of deciding who should be in one program or another, or who would be eligible to take a national test." [p. 2, 1991].

Future Trends. Dr. Kurt Geisinger position magnifies the trends among states toward performance-based testing, focusing on higher level skills and problem solving. States, that know the benefits to holistic and analytic scoring, have begun to explore assessment alternatives such as performance-based and portfolio assessment as an answer to the limitations of minimum competency testing for all disciplines and students. For example, Kentucky has discontinued norm-referenced testing in favor of a performance test for 1996. Louisiana has expressed an interest in performance-based assessment. Colorado and Connecticut have placed greater emphasis on performance and applications of student learning. Illinois expressed gradual increase in testing mathematics and higher level thinking skills. South Carolina is headed toward testing higher level thinking skills. Massachusetts will use proficiency scales in reporting an increase in the use of open-ended questions and will use performance assessment as supplements to the current program. Minnesota will add performance testing in science and health. Missouri will field test performance assessment in science.

Though these trends may reflect tests for some states on student competency through performance-based assessment alternatives, it is not necessarily as a function of curriculum mastery nor as partial criteria for a high school diploma. Schools are not headed toward deleting the mastery of a discipline as a criteria for a high school diploma.

Certainly, Texas is no exception with its testing program pressing "Onward Through the Testing Fog" in at least three directions. First, in October 1992 in grades 4, 6, 8, 10, a new norm-referenced test, Texas Test of Basic Skills [TTBS], will be administered in reading, writing, mathematics, science and social studies. The TTBS is being developed by Riverside Publishing for implementation to all students with the decisions on exemptions for limited English proficient students pending. Secondly, the TAAS will add disciplines of science and social studies by October 1993 in grades 5, 7, and 9. By October 1994, test publishers will have added grades 3 and 11 to these disciplines. And thirdly, a reanalysis of the entire testing program is proposed in 1993 by the Texas legislators for a report to Governor Ann Richards with hopes for a realistic answer to measure student learning outcomes.

2. **Assessing LEPs with Minimum Competency Testing** To test, when to test, or not to test LEP students? Dr. Kurt Geisinger found that states had varied testing practices for lim-

ited English proficient students. Table 1 identified this variation in at least 13 states using minimum competency testing as a criteria to issue a high diploma [Roeber, 1990; Valdez-Pierce, 1991].

The policy decisions used to determine eligibility exemptions for limited English proficient students can provide answers to the dilemma of "to test, when to test, or not to test." Among these 13 states decisions to exempt or not exempt limited English proficient students from a minimum competency test to receive a "standard" high school diploma, three patterns evolved.

Table 1

Exemptions or No Exemptions for LEP Students By States Who Link Minimum Competency Tests and a High School Diploma

States With Mandates	**Exemptions for LEPs**	**High School Diploma/ Passage of Test**	**Certificate of Completion**
Florida	Yes [1 to 2 years]	Yes	Yes
Georgia	Yes [Parent/District]	Yes	No
Louisiana	No	Yes	No
Maryland	No	Yes	Yes
Michigan	Yes [2 years]	Yes	No
Mississippi	Yes	Partial Exemption	No
Nevada	No	Yes	Yes
North Carolina	Yes	Yes	No
Ohio	Yes	Yes	Yes
Oklahoma	Yes [Parent/District]	Yes	Yes
South Carolina	No	Yes	Yes
Tennessee	No [1 year/District]	Yes	No
Texas	Yes [1 test/District]	Yes	No

First, a pattern called "sink or swim," seemed to be used by Louisiana which provided no exemptions and no optional certificates. Second, a "good neighbor" pattern seemed to be used by Maryland, Nevada, South Carolina, and Tennessee to provide no exemptions and offer a certificate of completion to recognize student differences. One difference became apparent with Tennessee offering an exemption only to the students who attended school in the United States for less than one year and not to students with limited English proficiency.

Then, a third pattern, "half-way," seemed to acknowledge individual differences based on language and academic abilities and offer eligibility criteria for students who can take the test with a degree of success. Nonmastery of the test could either mean no high school diploma such as in the Texas, North Carolina, Mississippi, Michigan, and Georgia or a certificate of completion such as in Florida and

Ohio. Curriculum alignment to the testing program and remediation is unclear for each state with the exception of Texas which does mandate a state core-curriculum and testing program.

For example, Texas mandates decentralized decisions on student exemptions from the Texas Assessment of Academic Skills, by having districts form a Language Proficiency Assessment Committee comprised of teachers, administrators, and/or parents, to determine eligibility for the first test administration. Texas offers unlimited test retakes until age 21 and requires remediation courses prior to subsequent test administrations. Now with a 9 percent increase of limited English proficient students in 1990 to 314,674 of which 13,000 were in grades 11 and 12, with Hispanics having the highest dropout rate, 44 percent in 1989, and with an increase in student enrollment to 60,000 from which the minority language groups represented the majority [i.e., Hispanics (74 percent); Asians 67,735 (6 percent); Native American 6,275 (3 percent)], policy makers are challenged to design assessment and curriculum alternatives to the limitations of state mandates (TEA, 1991a).

Mississippi provides a two year waiver and time for test retakes for the state's Functional Literacy Examination or the Subject Area Testing Program. Exemption is offered on the Basic Skills Assessment Program and the Stanford Achievement Test. A LEP Assessment Committee, consisting of teachers, testing coordinator, counselors, psychometric personnel, and principals, determines documentation for exemption. Guidelines with definitions on the different levels of English language proficiency are utilized with such assessment alternatives as reading inventories, writing samples, course grades, teacher observations, and tests [i.e., teacher-made, achievement, and language proficiency] to determine test eligibility.

North Carolina provides guidelines to differentiate language proficiency levels for test eligibility with consultation from an assessment committee. An exemption is offered when a student's English language level hinders test mastery.

A common linkage in the "half-way" pattern between Texas, Mississippi, and North Carolina seemed to be the recognition to define language differences, as suggested by Ed De Avila, and offering test retakes with the inference that remediation and time would ensure the student opportunities to master the test. The weight and penalty of one criteria as a decision factor for these students' success and productivity as citizens is still questionable at best. Though for some states, these three patterns offer answers to linking minimum competency testing and a high school diploma. They certainly present a narrow vision for student success and are not without penalty to the student.

One possible answer not addressed by Dr. Kurt Geisinger to demonstrate student learning outcomes and without penalizing the student may be found in the "individual pattern." Here, the district takes ownership to achieve student learning outcomes by concentrating on alternative assessment. California's Option 1 Alternatives uses this pattern by decentralizing the assessment of student learning outcomes and offering to provide districts with an opportunity to design an outcome-based assessment process to demonstrate educational results rather than one test to "clinch a high school diploma" as the indicator of competency (California Department of Education, 1991). Option 1 Alternatives, though not tied directly to a high school diploma, provides six alternatives to design an individual evaluation program to measure educational results of a district's student population. Table 2 delineates the option alternatives.

Table 2
California State Department of Education: Option 1 Alternatives

Alternative	Description	Group Tested
A: Comparable Achievement-Norm-Referenced Test	Employs a norm-referenced test to demonstrate performance at or above national average.	Reporting eligible LEP or Former LEP.
B: Comparable Achievement-CAP	Employs CAP scores to demonstrate performance at or above the state average.	Reporting eligible LEP at grade levels tested on CAP.
C: Gap Reduction-Norm-Referenced Test	Uses a norm-referenced test to establish that the gap is lessening between scores of LEP students and all students nationwide.	Reporting eligible LEP students.
D: Gap Reduction-CAP	Uses CAP scores to establish that the gap is lessening between scores of LEP student and all students statewide. [Actually successive cohorts CAP method].	Reporting eligible LEP students at grade levels tested on CAP.
E: Successive Cohorts GAP Reduction-Norm-Referenced Test	Uses norm-referenced tests to demonstrate improvement in academic achievement scores of successive cohorts of LEP students.	Successive cohorts of reporting-eligible LEP students at specified grade levels.
F: Design and Implementation of an Alternative Evaluation Method to Demonstrate Educational Results	Allows a district to design an alternative use of standardized tests or other assessment methods to establish that it is effectively serving its LEP students.	Variable-to-be-defined by approved study design.

BEST COPY AVAILABLE

From these option alternatives, Alternative F seems to provide more flexibility in designing assessment alternatives for a district's student population by suggesting to [1] exercise caution in attempting to the design of an alternative methodology, [2] to use only well-supported academic achievement data to document claims of academic parity, [3] to carefully document the validity and reliability of the selected evaluation design and instruments, [4] to base district evaluations on the broadest range of student achievement, [5] to set outcome standards high enough to ensure that LEP students really are academically successful, [6] to select achievement tests that match the district's curriculum and have appropriate difficulty levels, [7] to explain the educational principles on which the instructional program offered to LEP students is based, and [8] to analyze collected data using procedures that are appropriate for the hypotheses that are being tested and the research questions that are being asked (CDE, 1991).

Thus, individual patterns direct districts to become owners of their evaluation program, align curriculum to evaluation program, differentiate instructional program to diverse learners, and become responsible for student success to determine who graduates with a high school diploma. The state government then holds the districts accountable through reporting requirements for preestablished "real world" outcomes for their student population.

Limitations to each policy decision pattern are evident. Yet continued decisions on assessment alternatives, definitions on language and academic proficiency, alignment of curriculum to an assessment program can guide districts to make competent, consistent, and relevant decisions on the academic performance of limited English proficient students.

3. **Advantage or Disadvantage.** To extend Dr. Gesinger's view, the biggest disadvantage is expectations, respectively by teachers and districts. When state government holds districts accountable for student learning outcomes with one single measure, then districts reconsider their priorities in terms of the state government's expectations on educational outcomes. Teacher expectations are a disadvantage for the reasons that teachers expect less, teach to the test, teach less creatively, and differentiate learning opportunities. First, teachers with students in lower tracks generally receive less rigorous and lower quality instruction. Second, teaching to a test and to minimal skills often fragments concepts instead of treating topics in depth and involves rote memory instead of critical thinking. District expectations, in turn, reflect assignments of the more experienced and effective teachers disproportionately to higher tracks rather than to work with the lower achievers or students needing remediation.

Minimum competency testing creates, by its nature alone, minimum expectations which in turn affect student learning outcomes. In Texas, state test results still indicate after 10 years of minimum competency testing that students are not learning well and it certainly is not to say they cannot learn.

4. **Legal Issues.** Even though the courts have upheld the rights of states to use competency testing appropriately, the limitations placed by testing programs used in states certainly needs further investigation. First, Title VI of the Civil Rights Act of 1964 merits a reminder since it prohibits discrimination based on race, color, or national origin in programs receiving federal financial assistance. States and districts should be cognizant that this law pertains to the issue of equity and language minority students. States, which use minimum competency as an answer to determine school success, need to reevaluate their state standards and criteria in respects to discrimination of one's national origin. The importance is magnified when the development of testing standards do not coincide to student assignments and language abilities and thus place limitations on a student's academic success.

Second, the issue of tracking students on the basis of test results is also relevant. The landmark 1967 case of *Hobson v. Hansen* deserves investigation in which plaintiffs successfully changed the testing and tracking system that had been instituted in Washington, D.C. In this case, students were placed in three instructional tracks at the elementary level [i.e., special academic (retarded), general (average), and honors (gifted)] and added at fourth track at the high school level [above average (college bound)]. Each track provided a different curriculum commensurate with the students' tested abilities. The findings indicated that African-American students were placed disproportionately in lower tracks compared with the district's more affluent white students. The proposed national test certainly has chances to become another vehicle to "track" language minority students' successes and failures.

Third, another case, *Lau v. Nichols,* presented underpinning issues for school districts to take affirmative steps to rectify English language deficiencies. The case in Ann Arbor, Michigan, in 1979, has implications in which a federal judge ruled that Ann Arbor school must recognize that students who speak Black English may need special help in learning standard English. Black English may constitute a language barrier. If barriers do exist in attempting to teach standard English, then students may be feeling inferior. Teachers requiring students to switch from Black English to standard English impedes the learning of standard English. Thus, teachers would need to be trained on language differences and the impact in assessing standard English. Obviously, further investigation on the legalities of how testing impacts minority language groups' rights is warranted.

5. **Using Minimum Competency Tests with Limited English Proficient Students.** Dr. Kurt Geisinger's suggestions present limitations and merit discussion in the areas of remediation, LEPs for limited English proficient students, and curriculum alignment. First, I disagree that remediation after minimum competency testing to assist in mastery of the test objectives so students can receive a high school diploma. Remediation will assist a small percentage retakers at the high school level. The higher possibility exists that the retakers will not master the test objectives and dropout of school. For example, the preliminary Texas Assessment of Basic Skills test results in Grade 11 for April 1991 indicate 62,328 students were tested from which only 39 percent passed all three sections [mathematics, reading, and writing]; 4,860 students were retakers of which only 15 percent passed. Remember, retakers should have had the required remediation courses and counseling services. In addition, 12,628 students were retested with the TEAMS for a 57 percent mastery. TEAMS mastery does not indicate the number of test administrations by the retakers nor does it indicate the number of subtests taken.

In October 1990, 174,869 students were also given the TAAS in Grade 11 from which 65 percent mastered all three sections: Whites 75 percent, Asian 74 percent, American Indian 64 percent, Hispanic 52 percent and African American 45 percent. From the 13,659 identified limited English proficient students in Grades 11 and 12, 5724 were tested with only 18 percent passing, inclusive of retakers. Specifically, the discipline needing most remediation was mathematics with 74 percent mastery. Limited English proficient students, for obvious reasons, need remediation in all three disciplines with mastery as follows: reading 43 percent; writing 38 percent; and mathematics 40 percent. So when does one remediate?

Second, I disagree with the appropriation of monies and teachers to design specific individualized educational programs [IEPs] for limited English proficient students. These students' problems stem primarily from a lack of English language proficiency rather than a lack of academic knowledge. The isolated costs to districts would be monumental in the form of lower teacher-student ratio, evaluation experts and instruments, instructional materials, teacher training, and facilities to implement this solution. If districts were to assign IEPs for these students, then all students who have not mastered the test objectives should qualify. Instead of IEPs, the recommendations for alternative assessments to improve individualize instruction is suggested.

Third, I support the concept of curriculum alignment to testing. To change the process, districts must establish educational outcomes, think broadly, and align an assessment process with the curriculum and mastery standards. Remediation would then be integrated into the curriculum for each discipline as deemed necessary to master. Choosing the assessment alternatives would be dependent on the curriculum design and student population.

Two additional suggestions to Dr. Kurt Geisinger's position would include [1] to integrate teacher and policy decision training and [2] to develop an educational outcomes process. The first suggestion indicates the need for preservice and in-service training to successfully implement the instructional programs for limited English proficient students. Institutions of higher education, public schools, and state governments should actively collaborate to facilitate an understanding of assessment alternatives for limited English proficient students as presented by California. Today, teachers and policy decision makers [i.e., school board members, administrators, and counselors] have limited knowledge on how to assess students' academic progress much less their language proficiency, or how to interpret the test results to design and implement an aligned curriculum with an appropriate instructional program. Proper training for administrators and teachers is needed for policy to align itself with the instructional needs of limited English proficient students.

The second suggestion is the development of an educational outcomes process for limited English proficient students to include the definition of eligibility, alignment of curriculum and evaluation, integration of teacher training, integration of policy-decision training, identification of assessment alternatives, selection of program alternatives, and evaluation of program outcomes. The eligibility criteria for assessing educational outcomes would be the major first step to determine who should take or not take a competency test. This criteria could include a definition of language proficiency [i.e., listening and speaking] in English and the student's first language, a definition of language proficiency [i.e., reading and writing] in English and the student's first language, definition of academic proficiency [i.e., mathematics, science, social studies, health, fine arts, citizenship] in English and the student's first language, abilities on learning strategies [i.e., cognitive, metacognitive, affective, social], and the length of time in the school system.

Summary

The solutions suggested merit further discussion. Consideration of an educational process to access student learning outcomes is an alternative to the linkage of minimum competency testing for a high school diploma. Distribution of the ownership for student success will then be on the school system Therefore, if educators were to develop a consistent and equitable process to determine a limited English proficient student's educational outcomes, then that student will be guaranteed a quality education to be a successful and productive citizen in our society.

Note

[1] 269 F. Supp. 40 (D.D.C. 1967).

References

Bolte. G. (1990). Interview: Will Americans Work for $5 a Day? Time. 7- 23. (pp. 12-13).

California Department of Education. (1991). Option 1 alternatives: Technical standards and recommended practices for development of outcome based assessment of district services to limited English proficient students. Sacramento, CA: Author.

CTB Macmillan/McGraw-Hill. (1991). Defining limited English proficiency makes national test premature, expert tells subcommittee. CTB News Bulletin. (pp. 1-2).

Roeber, E.D. (1990). Survey of large-scale assessment programs: Fall, 1990. Lansing, MI: Michigan Department of Education, Michigan Educational Assessment Program.

Texas Education Agency. (1991). SnapShots '90: 1989-90 School District Profiles. Austin, TX: Author.

U.S. Department of Education. (1991). AMERICA 2000: An educational strategy. Washington, DC: Author.

Valdez-Pierce, L. (1991, March 26). Assessment issues concerning non-English language background and LEP students. Paper presented at the International Teachers of English to Speakers of Other Languages Conference, New York.

Reponse to Kurt Geisinger's Presentation

Lawrence M. Rudner
ERIC Clearinghouse on Tests,
Measurement and Evaluation

Minimum competency and graduation tests should be welcome by all. In theory, students lacking basic skills are identified and social promotions are ended -- thereby providing better quality and more appropriate educational opportunity. The flip side, passing the exam, should also be welcome. Passing serves as a certification that the examinee has achieved some level of competence and is ready for the next level of instruction or is worthy of being called a high school graduate. Professional standards and legal precedents hopefully assure that the tests are appropriate and that people are being classified fairly.

Unfortunately legal precedents and professional standards are not always followed when tests are used with Limited English Proficient students. The Children's English Services Study, for example, used a pathetically small sample to determine the score that was used to define students as Limited English Proficient (LEP). Most recently the U.S. Department of Education authorized the use of several inappropriate tests in order to have "something" for LEP students. The Secondary Level English Proficiency test, designed and validated to indicate whether a student has enough English skills to be mainstreamed into an English speaking classroom, for example, was approved as an admission tests for post secondary education. The Spanish version of the P.A.R. Ability to Benefit Test, based on the old Adult Proficiency Level (APL) examination, was also approved -- even though there were absolutely no statistics or documentation available for that version. It was merely a translation.

In his excellent review of minimum competency testing as it pertains to students with Limited English Proficiency, Geisinger (1991) describes the status of minimum competency testing, the methodological issues of MCT and relates those issues to the assessment of LEP students. Throughout his paper, Geisinger describes what I view most competent measurement specialists would advocate given the presented issue. He points to professional standards and does an excellent job of describing how they apply.

If followed uniformly, the guidelines and suggestions outlined by Geisinger would help assure fair and equitable testing. Indeed, Geisinger has worked as an expert witness testifying against companies that have not adhered to professional standards. The quality of assessing LEP students would be vaştly improved if all tests publishers and users followed Geisinger's recommendations.

Throughout my comments, I will be reiterating many of the goals espoused by Geisinger. We don't disagree on the goals and we don't disagree too much on what we view as the responsibilities of the profession. We might even agree on the theme of this editorial -- that commonly accepted practices don't go far enough to assure fair and equitable test usage.

The better test developers can point to numerous activities that they typically undertake to make tests fair and appropriate for all students, e.g. review panels, representation in norming groups, bias analysis. I will argue that while this current state of practice has positive effects, it does not sufficiently protect LEP students from being inappropriately labeled and classified.

I start by identifying a few key points made by Geisinger that I feel warrant further emphasis. I then discuss the concept of validity. With perfect validity, many of the MCT issues raised by Geisinger become moot. The issues are issues, however, because MCT tests, like all tests, are not perfectly valid. I discuss some of the steps described by Geisinger to resolve those issues, discussing why I don't feel they are good enough.

Points Warranting Reiteration

- We have a set of carefully drafted standards -- APA, Uniform Guidelines, the GRE guidelines -- that should be followed.

  These are statement by the profession outlining steps to assure quality assessment and meaningful documentation. The standards are rigorous -- even major test publishers typically fail to adhere to major standards.

- Test scores should not be used in insolation.

  On key advantage and hope for the current move toward "authentic assessment" is that multiple observations are used. A single test score is easily influenced by sampling error as well as individual variation. Multiple measurement has the potential to reduce this type of error. It is only a potential and not a given because authentic assessment has yet to clearly identify the universe of skills to which scores are supposed to generalize.

- Need to know basis information about policies concerning the testing to LEP students.

There has been very little systematic or in-depth analysis of testing policy and practices, let alone practices as they apply to LEP students. Such analysis is needed to initiate discussion. (The problem is not as bad as painted by Geisinger in the first draft of his paper. Finding no relevant articles, Geisinger's graduate student was evidently not proficient in searching the ERIC database. Our search yielded 30 article on Minimum Competency Testing and Limited English Speaking students and over 180 articles on testing Limited English Speaking students.)

- Testing should be educationally relevant.

  Diagnostic testing to help teachers identify weaknesses is much more useful than summative testing. One encouraging aspect of the current interest in testing and this conference is that educators, rather than statisticians, are taking control of testing activities. I fear, however, that the educators are being co-opted by the politicians.

- The effects (and use) of testing should play a major role in tests validation studies.

  While self-evident to some, this is considered a radical idea in the measurement community. If tests are to be used to promote the common good, then the social consequences of tests must be examined.

Validity

Geisinger cites the literature to identify several relevant validity concepts:

(1) we do not validate tests, rather, we validate the accuracy of inferences that we make from test scores (Cronbach, 1971)

(2) the degree of empirical relationship between test scores and criterion scores (Meassick, 1989)

(3) the extent to which a test may be said to measure a theoretical construct or trait (Anatasia, 1988)

(4) relevance of the content to the content of a particular behavioral domain of interest and about the representatives with which item or task content covers that domain (Messick, 1989)

(5) a measure of how well tests items represent the objectives of the curriculum (McClung, 1979).

The inference we want to make in a minimum competency test is whether the student has mastered some set of skills. The questions then are:

- do we have the right set of skills? and
- have we measured those skills adequately?

Let us assume for a moment affirmative answers to those two questions. If the skills are defined as those needed for success at the next higher level, then validity can be demonstrated empirically (definition 2). On a perfectly valid test, we would expect each tested skills to be a prerequisite for some higher level skill.

On the other hand, the skills on the tests may represent a theoretical construct -- for example, the skills a minimally competent high school graduate should have mastered (definitions 3, 4, and 5). On a perfectly valid test of this type, skills which should be mastered by the minimally competent appear on the test; skills that are not necessarily mastered are not on the test.

This right set of skills could be enormous. All the skills up to the minimal level should be included. If an individual fails to master a skill in the set, then that individual is not minimally competent. Many of the skills may appear to be trivial. Even if no minimally incompetent individual fails to master it, it belongs in the universe of skills mastered by the minimally competent.

If we have the right set of skills and all the skills are properly measured then English language skills don't matter. Either a student demonstrates minimum competency or he doesn't. Either he is ready for the next grade or not. Or, for a graduation examination, either he meets the definition or not.

Should the set "right skills" differ across population groups? Clearly not for a graduation exam; second class, standards are not equitable. Hopefully yes for a promotion exam. Hopefully our special programs for LEP students make a difference and have different prerequisite skills. As Geisinger points out, the curriculum for LEP students needs to be carefully examined. A well articulated instructional and testing program can greatly aid education. If it is poorly articulated, or if the relationship has not been examined, then the tested skills lack relevance, i.e., are not valid.

Standards and Adverse Impact

Close to the issue of "right skills" is the issue of standards. Tests typically have a passing score -- above which you are said to be competent, below which you are not. The need for passing scores is an admission that we may not have "right" skills. If you need to be minimally competent to pass an item, then 100 percent of the minimally competent would get the item right and the passing score would be 100 percent. Herein lies problem number 1, we are not very adept at defining domains. We include skills that minimally competent people get wrong. We admit as much when applying a standard setting technique such as the one attributed to Angoff and ask "What proportion of minimally competent people will get this item right?" Tests are not perfectly valid, and we don't have any ironclad standards.

It would be nice if we had a test which measured the right skills and had an incontrovertible standard above which everyone is competent and below which everyone is incompetent. Adverse impact would not be an issue.

Adverse impact occurs when members of one group are underepresented by the selection rule. With top-down selection, for example, individuals are selected based on their ranking, starting with the highest score and working downward until all available slots are filled. If the group means are different, then the members of the group with the higher mean will be selected first and will occupy most of the available slots. The de facto standard is well above the minimal competency level. While capable, members of the lower scoring group are systematically denied access. There are numerous court cases, *Griggs v Duke Power* being the most famous, where employers intentionally discriminated under the guise of an objective test. (While Geisinger did an excellent job of describing the *Debra P* case, which was an MCT case, he did not describe an entire body of legal precedents which I know he knows well.)

Tests cannot be used to exclude systematically if everyone is simply rated as competent or not. Those that are doing the selecting must choose from a pool of qualified applicants. If people are randomly selected, that is, given equal opportunity, then the proportion of group members that are selected would be the same as the proportion of group members that are qualified. Of course, MCTs are rarely used just to make a dichotomous classification. Scores are used and individuals are ranked. We don't have pure MCTs.

Standards and Adjustments

Geisinger provides a list of nine pieces of information that may be used to adjust the standards on the kinds of MCTs that are usually developed. In my own research on standards for teacher licensure examinations (a form of MCT), I found that the standards were lowered to ridiculous levels -- to the point that the tests were meaningless.

The usual argument is to adjust for false negatives -- failing people that are really competent. While false negatives may scream louder, they are rarely more serious of a consequences than are false positive -- certifying someone who is not really competent. Indeed, I will argue for an upward adjustment: I would rather leave someone back than to incorrectly promote them. Neither should be given preference.

The false negative argument is closely allied with the adverse impact argument -- "we lowered the standards so more LEP students would pass." While such an action may be politically astute, it is not educationally sensitive. Promoting people that are not qualified obviates the entire purpose of the testing program.

Geisinger argues for adjustments in the name of measurement error due to unreliability. He advocates lowering standards if the reliability is lower and the standard error of measurement is higher for LEP students. If the reliability is that different, then perhaps the test should not be used. Downward adjustments are not justified as measurement error can be in either direction. Finally, one would expect different reliability estimates simply due to variance differences. The act of making adjustments, however, is an admission that there is either a problem with the test or with the standards as they stand. They are not necessarily "minimum" standards.

Using Group Data

To help assure that we have measured skills adequately, the better test developers make sure groups are adequately represented in the item tryout and norming studies. Group data such as this, however, can easily mask real differences. Suppose we have a norming group for a mathematics test made up of 80 percent English skills, and 20 percent LEP students, a 5-option multiple-choice item; the item p-value is .60 for native speakers; the item p-value is .40 for LEP students with adequate English skills, and the item's requisite English skill is a problem for 25 percent of the LEP students.

The fact that the English load is a problem for 25 percent of the LEP students should raise some flags. When LEP students get the item wrong, we don't know if it is because they legitimately don't have the math skill or if their lack of English caused the problem.

If there were no English load, the p-value for this item from the norming study would be .56 (8*.6 + .2*.4). The inappropriate English load lowers the p-value to.55 (the LEP contributions is .75*.2*.4 for students with adequate skills plus .25*.2*.2 for students with inadequate skills since they can guess). The inclusion of LEP students would have no appreciable effect on the norms or item statistics.

We would expect a bias analysis to flag an item that presents an inappropriate problem for 25 percent of a population. Using the above example, the LEP p-value for LEP students is .35 (.75*.4 + .25*.2). We would compare this to the expected value, .40, and conclude that there is no bias. The problem stems from using a heterogeneous group, LEP, to look for problems that occur with only some members of the group.

Recommendations

Recognizing that we have to use tests and standards that are not infallible, I would rather see the same standard for everyone and measures of the goodness-of-fit (individual assessment accuracy, person-fit) calculated. A goodness-of-fit could simply be the correlation between an individual's response pattern and the item difficulties. We expect people to get the easy items right and the hard items wrong. If an individual's response pattern, regardless of English skill or race, doesn't make sense then the test data should not be used. Testing problems should be identified at the individual, not the group, level.

LEP students need to be included in norming studies; bias analysis needs to be conducted; standard setting studies need to be conducted. These steps outlined by Geisinger will improve norms, identify many flagrantly bad items, and help establish meaningful standards. Following these steps will clearly improve the quality if not the credibility of a testing program. If we are interested in developing assessments that are truly applicable to all children, LEP and non-LEP, then we need to do a better job of identifying the skills that we want to assess and a better job at identifying which students were properly assessed and which ones were not.

Innovative Practices in the Identification of LEP Students

JoAnn Canales
North Texas State University

The development of the paper was rather like *deja vu*. Since 1974, when I entered the public school system as a recent graduate of a Speech Pathology program, I had all the answers...until I started working with children in a Chapter 1 identified campus in a border town school district. Each year, teachers would refer entire classes to me, and I realized quickly that I did not have the slightest idea how to tell the difference between those in need of Speech Pathology services and those in need of English language development services.

Two decades later, we still wrestle with the same issues, and I submit to you that, given the background of the students now entering the public school system, more and more students will be in need of English language development/Speech Pathology services related to articulation and language disorders, regardless of their ethnic or linguistic background.

Thus, the purpose of this paper is to describe current practices in various states used to identify linguistically different students, provide a review of the literature regarding recommended practices, and offer alternative practices for identifying linguistically different students. The expectation is that the information contained herein can serve multi-fold purposes:

1. provide an information base regarding current identification practices
2. suggest a way to systematically identify limited English proficient students using multiple criteria; and
3. offer a paradigm that will allow the United States Department of Education and the various state departments of education to collect consistent data regarding the students in need of English language assistance.

Methodology

To this end, in addition to a review of the literature, surveys were mailed to 17 states that provided a geographical representation of the eastern, heartland, and western regions as well as a multilingual and multicultural representation of the 17 states surveyed, 9 responded. These states graciously responded within a two-week time frame which is most deeply appreciated and acknowledged.

The recommendations in the section entitled "Paradigm for Determining English Language Assessment Needs" seeks to incorporate yet expand current practices extant in the various states. The intent is to make the modification of traditional practices more palatable and pragmatic which will enable practitioners to move toward the use of multiple criteria for identification and assessment of linguistically different students.

Review of Language Assessment Practices in Selected States

The purpose of the survey was to obtain data on the LEP population and the English speaking population by grade level with respect to ethnicity, language(s) spoken, and program offerings and to examine these data for any relational patterns between the size and the type of the LEP population versus the identification and assessment practices in the various states.

The limited information received as a result of the survey precluded making any generalizable observations. An attempt to utilize data provided by another national study (Olsen, 1991) yielded some discrepancies between data provided in the report and data provided by some of the states surveyed. Thus, efforts to address the intent of the survey were not very successful.

Sufficient information was provided, however, regarding the identification and assessment practices utilized to make the following observations:

Home Language Surveys (HLS) are used by each of the responding states as the initial screening instrument although the number of items on the HLS varied from state to state. Also, some states, such as New Mexico, use ethnicity as the identification criteria on the HLS and others use languages spoken. Variations in these instruments generate different kinds of information that can be collected regarding LEP populations. One additional factor that may be problematic in using this self-report type of instrument stems from misinformed parents or guardians who feel a need to misrepresent the native language spoken in the home. Such parents often feel that their children will be placed in programs that are not conducive to learning English if they respond truthfully on the HLS.

Standardized Achievement Tests (SATs) are used by every state, however, the cutoff score for identification, and exit criteria, varies between the 23rd percentile and the 40th percentile. This large discrepancy between cutoff scores will significantly impact on the number of LEP students identified per state.

Oral Language Proficiency Tests (OLPTs) are also used by every state although some states, such as New Mexico, limit their recommendations to four specifically listed OLPTs and others, such as Texas, list eight possible options. Inter- and intra-state variations in the OLPTs utilized also contribute to inconsistent identification and data collection practices because there is no correlation between the various instruments.

Some of the states suggest the use of optional criteria and merely list the possibilities, e.g., interviews, observations, and classroom performance, while other states (Louisiana, New Mexico) suggest specific interview techniques or checklists for specific performance behaviors. Regardless of the optional criteria used, the difficulty lies in that there is no apparent means of correlating performance on these alternative measures with their performance on the SATs or the OLPTs.

Additionally, many states allow each school district total autonomy regarding procedures utilized. This factor, coupled with the wide variation in practices, has implications for collecting consistent data regarding the number of LEP students, the kinds of languages spoken, and the level of assistance needed. Further, it makes it extremely difficult to conduct statewide or nationwide research on programs serving LEP students that will yield consistent, credible, and defensible data for decision makers in the field.

Recommended Integrative Approaches to Language Assessment

In reviewing the states' practices for identifying LEP students, two criteria surfaced repeatedly as being used extensively, although the manner in which these criteria were used varied. These two criteria are the standardized achievement tests and the oral language proficiency tests. Much has been written about the inadequacies of standardized achievement tests and oral language proficiency tests as measures of an individual's proficiency in English (Canales, 1990; TEA, 1988; Oller, 1973). Regardless of their shortcomings, to date, they have been widely used by the majority of the states as a basis for consistent measurement of students' linguistic performance. Since the 1970s, however, several options have been recommended that would provide practitioners with a more realistic and comprehensive assessment of an individual's English language proficiency (Canales, 1990; Erickson, 1981; Thonis, 1980; Oller, 1973). Some states reported using these measures, or at least recommending them as optional measures in their state publications.

These optional measures assess language proficiency while a student is engaged in a meaningful speech event. This is known as an *integrative* approach to language assessment because students utilize several communication skills simultaneously. The use of these recommended measures to assess an individual's *integrative* use of language skills is necessary because, heretofore, primary measures of language assessment, namely SATs and OLPTs, have focused on discreet items of language proficiency, e.g, use of verb tense, use of correct vocabulary term. This process severely limits the amount of information regarding an individual's actual proficiency with a language because language usage:

1. is dynamic and contextually based (varies depending upon the situation, the speakers, and the topic)

2. is discursive (requires connected speech)

3. requires the use of integrative skills to achieve communicative competence.

This definition of language usage is predicated on a socio-linguistic theoretical base suggesting that language is more than just a sum of its discrete parts. The implication then is that language assessment instruments also need to follow a similar theoretical base, a practice that has historically been ignored in traditional language assessment procedures (Canales, 1990).

Language assessment instruments consistent with this philosophy are known as measures of integrative skills and include observation instruments (rating scales and checklists), interviews, dictation tests, and cloze instruments. A description of each follows.

Observation Instruments

Classroom observations of students interacting in various settings are the basis for determining students' linguistic proficiency. A student's linguistic performance in listening **and** speaking is rated on a five-point scale of proficiency, ranging from non-native speaker of English to proficient speaker of English, for each of the four linguistic subsystems -- graphophonemic (letters/sounds), lexicon (vocabulary), morphology (grammar), and semantics (syntax/meaning) (see Appendix A & B). These rating scales are completed by the classroom teacher after observing students in various classroom settings. Separate rating scales can also be completed for observations of casual, social interactions, such as playground or cafeteria talk. Appropriate completion of these rating scales requires that the classroom teacher have an understanding of the criteria used to rate each of the linguistic subsystems.

The behaviors on the rating scale can also be listed in a checklist format in increasing order of difficulty for ease in scoring and analysis.

Interviews

Structured interviews are developed and administered on an individual basis. Ideally, an examiner should conduct the interview while a language specialist transcribes the examinee's responses, noting the use of the four linguistic subsystems. The advantages of this kind of measure are that it can be individually tailored to the experiences of the examinee and it allows the examiner opportunities to explore an individual's knowledge of the language.

The disadvantages, however, are several. First, it usually requires two people to administer the interview, a skilled interviewer and a language specialist. Second, this interview scenario has the potential to distract the examinee and perhaps contribute to diminished responses because of intimidation, especially for young children. Third, individualized administration makes it a time-consuming procedure. Finally, without appropriate scaling criteria, interviews are unsuitable for widespread use in schools as a tool for identification and placement of students.

Dictation Tests

The examinee listens to text dictated from graded material and writes down what is heard. The premise for this measure of integrative skills is that the individual needs to have knowledge of the four linguistic subsystems in order to convert speech to print. The use of dictation tests is advantageous because they:

- are easily developed from material used in everyday classroom situations such as basal readers, science books, or social studies books;
- can be administered in a group setting; and
- do not require extensive specialized training to develop or administer.

The few disadvantages of dictation tests, which can occur in the administration phase and the scoring phase, are manageable if the examiner is aware of them. First, an examiner's dialectal differences may cause difficulties in transcribing speech to print, a problem that could be overcome by using a taped version of the dictation. A related problem, students' lack of familiarity with this type of test, can be mitigated with practice sessions prior to the **actual** dictation to be used as the measure of language proficiency.

Second, an examinee's unfamiliarity with all of the variations in spelling of English sounds may cause interference for the examinee in converting speech to print, for example, writing "miss is esmith" for "Mrs. Smith," for example. This difficulty can be overcome by having the dictation tests scored by someone who knows the differences between the graphic and phonetic systems of the examinee's native language compared to the system in English.

Third, the dictation test requires that the individual being tested knows how to write and finally, appropriate criteria for scaling need to be developed as in the case of the interviews.

Cloze Instruments

The examinee is asked to complete a readability-graded passage from which words have been omitted at regular intervals (usually every fifth word). The premise of this procedure is that language is highly redundant, with many contextual clues that can inform the examinee of the appropriate missing words if that person has a command of the language being tested. Cloze instruments have been used for many years and validated by reading specialists. Administered and analyzed properly, the results of cloze tests will yield information regarding the examinee's level of facility with the text. Such information is useful in planning for students' instructional needs.

In addition to its instructional orientation, there are many advantages to this procedure. The test can be prepared easily using texts that students use in the classroom, thus making the assessment procedure a functional one. Further, the test can be administered in a group setting and quickly scored. If administered to native English speakers at the same grade level, their scores can serve as a basis of comparison for the non-native speakers' scores. Additionally, the construction, administration, and scoring of the cloze test do not require any extensive specialized training to use correctly.

The difficulty in implementing the use of integrative measures of English language proficiency lies in the lack of

- broad based acceptance with respect to their ease of development and administration,
- understanding of the breadth and depth of their usefulness, and
- standardized procedures for consistently collecting and correlating alternative data on students.

These factors preclude the use of 'integrative measure' data in making uniform decisions regarding the identification, placement, and exit needs of LEP students.

Following is a model for ameliorating this dilemma. The scope of the model, however, exceeds the traditional practice of identification and can be used to make decisions for placement and exit, as well. Use of this model consolidates the gathering of information for practitioners and enables them to make informed decisions regarding the needs of the linguistically different children.

Paradigm for Determining English Language Assistance Needs

The model mentioned above is a comprehensive process that identifies not only students in need of language assistance but the level of assistance needed as well. The process involves a systematic documentation of students' linguistic proficiency in formal and informal settings and academic and non-academic settings. In short, this process generates a profile of a student's needs for language assistance and thus, has been titled the English Language Assistance Needs (ELAN) Profile Chart. The ELAN Profile Chart enables practitioners to document data needed to appropriately meet the instructional needs of students and the programmatic needs of campuses.

There are specific steps that must be addressed prior to implementing the effective use of such a model. These steps include

- identifying **criteria** to be used,

- developing a **Likert rating** scale to accompany each criterion,

- determining the range of scores possible for each **category of need,** and

- designing and implementing the **training** necessary to institutionalize the process.

Specifically, each step entails the following considerations.

Criteria Development

A comprehensive assessment of a student's language assistance need(s) requires that data be gathered in three areas. These three sets of data include non-academic related **o**ral language proficiency data, **s**ocial data, and **a**cademic data (**OSA**). In each of these areas local/state education agencies have the flexibility to include as many options as are feasible to be undertaken. The important consideration is that each option be clearly delineated and available to all of the individuals involved to ensure consistency of implementation. Some of the examples of the types of options have been mentioned in the section entitled "Review of Language Assessment Practices in Selected States" and discussed in the section entitled "Recommended Integrative Approaches to Language Assessment." These options, and others, are listed below along with a brief rationale for their utilization.

Oral Language Proficiency Data

Home Language Survey -- This serves as an initial screening and is currently used in many states. It can provide useful information regarding baseline data such as language(s) spoken in the home.

Oral Language Proficiency Test -- These prepackaged instruments provide inexperienced practitioners with baseline data regarding students' linguistic performance albeit minimal data.

Oral Language Interview Instruments -- These instruments enable interviewers to probe for information not readily accessible through pen and paper tests.

Observation Instruments -- Provide detailed, comprehensive data on students as they engage in actual speech events which minimize the intimidation factor present in other testing situations.

Social Data

Socio-Economic Status (SES) -- An often disregarded criterion, the SES of a student can offer valuable information regarding the amount of oral/aural stimulation received in the home. Typically, children from low to mid SES home environments are not likely to have

- engaged in much dialogue,
- been read to by their parents,
- or experienced summer camps, organized sports, or other similar experiences that help develop linguistic skills.

Schooling Experience -- This, too, is an often disregarded criterion. Information gained can inform practitioners about the possible level of skills learned in a formal school setting. If these skills are not continuously developed or are developed in a country other than that of the target community, students will need additional intervention services.

Observation Data -- This information obtained from the home and other social settings such as the playground, the cafeteria, etcetera can validate, or confirm, other data gathered.

Academic Data

Achievement Test -- Standardized achievement tests have been a primary source of data used by many states. As mentioned previously, however, the cut off score for eligibility has varied from state to state. Many states also use state-specific standardized tests. Unless these instruments are administered at each grade level, such instruments will not provide consistent data and, thus, are not recommended for use as criteria.

Cloze Test -- Used by many states, such instruments provide useful data regarding the students' language proficiency level with classroom text information that is the basis for participation and promotion in the schooling process. Its ease of administration and scoring make it a valuable criterion for consideration.

Six Weeks Grades -- This criterion provides formative data on students' performance and is the primary criterion used for promotion. The mean should be monitored during each six weeks across subject areas and the mean for the first five of the six weeks should be used as one of the criteria for assessing English language assistance needs. Individual school agencies need to establish specific subject areas to include in the mean.

Observation Data -- Checklists or rating scales utilizing specific performance criteria can provide information regarding students' use of language in contextual situations.

While the number of criteria suggested above may seem unreasonable, multiple data are necessary to develop a consistent and defensible process for documenting the identification, placement, and progress of LEP students and the benefit of effective programs needed to serve them.

Likert Rating Scale Development

The second necessary step in the process is the development of a rating scale for each criterion to be included in the ELAN Profile Chart (see Appendix O). A five-point scale is recommended to provide consistency across sites using a similar procedure. Following are examples of suggested scales as well as brief rationales/explanations for the descriptors accompanying each rating.

Home Language Survey

1 -- Only Native Language Spoken
2 -- Mostly Native Language Spoken
3 -- Native and English Languages Spoken
4 -- Mostly English Spoken
5 -- Only English Spoken

Most of the home language surveys presently used by state or local education agencies ask three to eight questions that would yield this information. Examples of some of the questions include,

- Which language did your child first learn to speak?
- What language does your child use most often at home?
- What language do you most often use to speak to your child?
- What language does the father speak to his child most of the time?
- What language does the child speak to his/her father most of the time?
- What language does the mother speak to her child most of the time?
- What language does the child speak to his/her mother most of the time?
- What language does your child speak to his/her brothers and sisters most of the time?
- What language does your child speak to his/her friends most of the time?

Oral Language Proficiency Instrument

1 -- Non-English Speaker
2 -- Extremely Limited English Proficiency
3 -- Limited English Proficiency
4 -- Near Native-Like English Proficiency
5 -- Fluent English, Native-Like Proficiency

The descriptors for this scale reflect those found in OLPTs adopted for state use. Each descriptor has a range of possible scores based on the students' performance on the test.

Oral Language Interview Instrument

1 -- 80-100 percent Native Language Responses
2 -- 50- 79 percent Native Language Responses
3 -- < 50 percent in either Language
4 -- 50- 79 percent English Language Responses
5 -- 80-100 percent English Language Responses

This scale can be applied to any interview instrument regardless of the number of items contained therein. While specific response criteria is not provided, the expectation is that the interviewer will have been appropriately trained to score acceptable responses.

Observation Data

1 -- Pre-Production Stage
2 -- Early Production Stage
3 -- Speech Emergence Stage
4 -- Intermediate Stage
5 -- Fluent Stage

These are widely used labels for the various stages of language development (references). Specific behaviors relevant to each of the stages can be found in Appendix C.

Socio-Economic Status

1 -- < $5,000
2 -- $5,000 - 10,000
3 -- 10,000 - 25,000
4 -- 25,000 - 35,000
5 -- 35,000 - 45,000

These ranges are partially arbitrarily based on the qualifications for free and reduced lunch as well as a general approximation of the relative cost of meeting the basic needs of a family versus the affordability of "frills."

> [Note: Perhaps a more precise scale can be determined using the current Poverty Level Index that considers the number of family members versus the income.]

Schooling Experience

1 -- No Previous Schooling or All English Program Only
2 -- Interrupted Schooling/Some ESL Instruction
3 -- Schooling in Other Countries
4 -- ESL program only since entering U.S. school system
5 -- Bilingual education program only since entering U.S. school system

This factor is critical to successful participation in the academic setting. Students with little or no previous formal schooling experiences or students placed in inappropriate programs will be in need of extensive linguistic and cultural education services.

Observation Data (Home, with friends)

1 -- Uses native language ONLY in all settings
2 -- Relies on native language in all settings
3 -- Uses the native language sparingly in all settings
4 -- Uses the English language with friends only
5 -- Uses the English language mostly in all settings

Knowledge of language use in various settings can also indicate the possible level of proficiency with respect to vocabulary development.

Standardized Achievement Data

1 -- < 20 %ile
2 -- 20-29 %ile
3 -- 30-40 %ile
4 -- 41-59 %ile
5 -- 60-80 %ile

The distribution of percentile points for each rating decreases from 20 to 9 because of the critical need to have a command of the language in order to perform well on these tests, recognizing of course that knowledge of the English language is not the only critical factor central to performing well on these measures. It should be noted that the ratings of 1 and 2 exceed the maximum cut-off scores found in states with large populations of linguistically different students, however, this type of scale can provide consistency in identification data and is thus presented as such.

Cloze Test

1 -- Raw Score of 0 - 20
2 -- Raw Score of 21 - 30
3 -- Raw Score of 31 - 40
4 -- Raw Score of 41 - 49
5 -- Raw Score of 50

Cloze measures can be statewide versions based on state adopted texts or local versions. Decisions will need to be made regarding which content areas to include as cloze texts.

Six Weeks Grades

1 -- <= 59
2 -- 60's
3 -- 70's
4 -- 80's
5 -- 90's

The six weeks grades for each of the content areas can be used as a formative measure to monitor additional needs for English language assistance. The mean of the six weeks grades for the first five six weeks, either for individual subject areas or across subject areas, is recommended to assist decision makers in the early identification of students in need of English language assistance for the subsequent school year. Subject areas to be considered for determining this mean should at least include Language Arts, Science, and Social Studies given the language demands of the disciplines.

Observation Data by Grade Level and Subject Area

1 -- Points, identifies
2 -- Names, lists
3 -- Describes, tells (simply)
4 -- Compares, describes (more complex)
5 -- Analyzes, synthesizes

Linguistic information obtained as students engage in academic work can be particularly insightful for making programmatic decisions for these students. This information can be obtained using checklists or rating instruments once the desired behaviors have been identified (see Appendices D-L).

The ratings for each criterion presented above can easily be recorded in sample charts provided in the Appendix section of this paper. Appendix M illustrates an **Individual** English Language Assistance Needs Profile Chart and Appendix N illustrates a **Campus** Language Assistance Needs Profile Chart for use in recording the pertinent data.

In some instances, decisions will need to be made regarding missing data or non-applicable data. Suggested for use are "M" for data that is **M**issing and "0" for data that is not applicable, so that it will not get factored into the total count. Comments about why the descriptors were not applicable would be helpful in informing future users of the data and alerting them to changes which may need to be made. This procedure will ensure consistency in and utility of data collected.

Distribution of Scores by Category of Need

Once the criteria and the ratings have been determined, the next step involves the distribution of the number of points possible into each of the categories of needs -- Beginning, Intermediate, Advanced. Given the descriptors attached to each rating, the greater the number of points accumulated per child, the greater the child's proficiency in the English language. In contrast, the fewer the number of points accumulated for each child, the greater the demand for English language assistance. This inverse relationship between points accumulated versus need is consistent with current practices in the various states. Such that, if students are at a "Level 3," they are at the advanced, near proficiency stage, and if they are at a "Level 1," their proficiency in English is virtually non-existent.

To further illustrate this point, if 11 criteria are selected to include in the ELAN Profile Chart as suggested above, then the greatest number of points would equal 55 [5 (rating) x 11 (criteria)] and the least number of points possible would equal 11 [1 (rating) x 11 (criteria)]. An individual student can total less than 11 points if there are some data that are Not Applicable (see Note below). An example, of the distribution of points is provided below.

34 - 55	Advanced Stage (Total possible if student scores all 5s or some 5s & 4s)
23 - 33	Intermediate Stage (Total possible if student scores all 3s or some 3s and 2s)
00 - 22	Beginning Stage (Total possible if student scores all 2s or 1s)

[NOTE: A score of 0-10 might be possible if there were missing data. If the criterion was important enough to include, decision makers may want to monitor the student's performance until the necessary additional information is available.]

As with every process conceptualized for wide use, certain realities, such as lack of resources, often preclude the comprehensive and extensive use of recommended procedures. In those instances, the following alternative is offered:

1. Deduct five points per criterion omitted from the overall total and adjust the totals in the three categorical levels accordingly.

55	Total in example (11 criteria)
- 5	Oral language interview
- 5	Observation Data (Social)
45	New Total for 9 criteria

28 - 45	Advanced (Scored all 5s or some 5s & 4s)
19 - 27	Intermediate (Scored all 3s or some 3s and 2s)
00 - 18	Beginning (Scored all 2s or 1s, and possibly some 0s)

2. Add five points for each criterion included to the overall total and adjust the three categorical levels accordingly.

55	--	Total in example (11 criteria)
+ 5	--	State-wide test administered at each grade level
60	--	New Total for 12 criteria

37 - 60	Advanced (Scored all 5s or some 5s, 4s, & 3s)
25 - 36	Intermediate (Scored all 3s or some 3s and 2s)
00 - 24	Beginning (Scored all 2s or 1s, and possibly some 0s)

If the school records of students are unavailable due to high mobility factors or recent immigrant status, then certain criteria may be selected in order to identify language assistance needs upon the student's arrival. For example, the Home Language Survey, the Oral Language Proficiency Test, the Previous Schooling, and the Oral Interview data can all be obtained readily. The distribution of scores would then be adjusted accordingly so that decisions regarding need and placement could be made. This would ensure that the student received appropriate services pending the arrival or attainment of additional information such as SAT scores or grades.

Advantages of the ELAN Profile Chart

Although at first glance, the process may seem cumbersome, the ELAN Profile Chart has many potential advantages. Some of these advantages include:

Teacher judgment is systematically documented.

Comprehensive information regarding a student's language proficiency is uniformly documented and available for use by teachers or parents.

Needs assessment can be conducted during end of the year LPAC meetings which, in turn, can facilitate student and faculty assignments for successive years.

Consistency in the identification process is possible in that the categorization of English Assistance Needs levels are based on Likert scale totals with corresponding points of distribution regardless of the number of criterion used.

Autonomy and flexibility in the criteria to be utilized remain a viable option for the state and local education agencies yet enable the United States Department of Education and the state education agencies, respectively, to collect data on the number of students in need of language assistance.

Identification, placement, and exit criteria systematically documented enable Language Proficiency Assessment Committees to execute their responsibilities conscientiously, consistently, and equitably.

Paper work is reduced to a manageable level, utilizing the comprehensive ELAN Individual Profile Chart (see Appendix M) or the ELAN Campus Profile Chart (see Appendix N).

Future Directions

Four critical mega-steps, if you will, need to be accomplished in order to implement the use of an ELAN Profile Chart.

First, the criteria to be utilized must be determined, or developed as in the case of the observation instruments. **Second**, participants in the process will require training in the development and usage of the instruments. **Third**, the data collected annually should be evaluated quantitatively and qualitatively to assess any patterns and note any anomalies. **Fourth**, longitudinal data should be cross validated for accuracy so that adjustments in the Likert scales can be made accordingly.

WHEATLAND ELEMENTARY SCHOOL DISTRICT

WHEATLAND ORAL LANGUAGE ASSESSMENT FOR DIAGNOSIS AND PLANNING

RECEPTIVE LANGUAGE - LISTENING

NAME__________________ GRADE_______SCHOOL YEAR 19__ - 19__

SCHOOL_________________ Context: __ Home __ Peers __ School Primary Language________________

Indicate the pupil's language competencies by circling the appropriate rating. Use RED in October; BLACK in May

---

	Phonology The Sound System		Vocabulary The Lexical System		Grammar/Syntax The Structural System		Semantics The Meaning System
5	Understands immediately all speech in any formal classroom or informal presentation	5	Understands words at a very advanced level within context of formal classroom instruction and informal situations	5	Understands large units of connected discourse presented rapidly and grasps easily necessary grammatical and syntactical cues	5	Understands all classroom instruction and informal conversation very well; obtains meaning from oral context immediately
4	Understands classroom instruction and informal speech at the normal rate of presentation	4	Understands words beyond the basic level necessary to manage classroom instruction	4	Understands complex sentences containing several ideas and given at the normal rate; has ordinary grasp of syntax and grammar	4	Understands classroom instruction and informal conversation and obtains meaning from spoken context; occasionally asks for repetition of materials
3	Understands most of what is said, but occasionally asks for repetition	3	Understands words basic to the formal instructional program	3	Can understand simple sentences of moderate length when given at a slow rate; usually grasps word order and grammatical cues	3	Understands ordinary classroom instruction and comprehends the context or daily activities at the level of an average native speaker
2	Can understand what is said if speech is slowly and carefully pronounced	2	Understands only those words related to simple classroom instructions and informal directions	2	Can understand very short, simple sentences, given at an abnormally slow rate of speed; has some difficulty with syntax and grammar	2	Understands some of what is said in class and comprehends the context of only very simple material when slowly presented
1	Has great difficulty understanding what is said and appears to confuse sounds	1	Understands only those words necessary to basic needs	1	Has great difficulty in understanding more than one word statements	1	Understands relatively little of what is said in class and does not obtain meaning from the context of orally presented material

Appendix B

WHEATLAND ELEMENTARY SCHOOL DISTRICT

WHEATLAND ORAL LANGUAGE ASSESSMENT FOR DIAGNOSIS AND PLANNING
EXPRESSIVE LANGUAGE - SPEAKING FLUENCY

NAME__________________ GRADE________SCHOOL YEAR 19__ - 19__

SCHOOL______________ Context: __ Home __ Peers __ School Primary Language____________

Indicate the pupil's language competencies by circling the appropriate rating. Use RED in October; BLACK in May

	Phonology The Sound System	Vocabulary The Lexical System	Grammar/Syntax The Structural System	Semantics The Meaning System
5	Pronounces words and sentences without error and with the fluency of a native speaker of commensurate age	Has an excellent vocabulary; uses figurative language and idiomatic expressions with the ease and facility of a highly fluent native speaker	Speaks very fluently in the classroom and in informal settings; oral expression is comparable to the language proficiency of native speaker of commensurate age	Speaks in a totally comprehensible manner in any situation; fluency is that of a highly verbal native speaker
4	Pronounces words and sentences at a level comparable to the language skills of an average native speaker	Has an above average vocabulary; may occasionally use figurative language and idiomatic expressions of a native speaker	Speaks in a generally fluent manner in the classroom and in informal settings but may demonstrate occasional difficulty in expression or the influence of non-native speech	Speaks in a comprehensible manner; functions well in classroom and informal context with no errors or difficulty; performs at the level of an above average native speaker
3	Pronounces words and sentences in a manner that can be understood but there is evidence of occasional error or influence of non-native speech	Has a vocabulary estimated at an average level for a native speaker and can manage adequately the subject matter idioms and figurative language	Speaks fluently and can express ideas adequately in the classroom and other settings; language competency appears to be that of an average native speaker	Speaks in a reasonably understandable manner like that of an average native speaker and appears to make sense most of the time
2	Pronounces words and sentences which are not fully understood because of errors which may cause comprehension problems	Has a limited vocabulary; may experience difficulties in expressing ideas; does not use idiomatic or figurative language	Speaks haltingly and has some difficulty expressing ideas; makes many grammatical errors and is frequently unable to respond	Speaks in a manner which often cannot be clearly understood; does not appear to be expressing self in a comprehensible way
1	Pronounces words and sentences which cannot be understood at all	Vocabulary deficits result in the wrong use of words, idioms or figurative language; usually silent	Often silent; rarely speaks at all except in a very limited manner in informal settings only	Cannot express meaning in connected discourse at all; expressions are usually one or two words sometimes accompanied by gestures

BEST COPY AVAILABLE

Identifying & Instructing the LEP Student
(English as a Second Language)

Levels of Proficiency	Level I	Level II	Level III	Level IV	Level V
ESL Categories	Pre-Production Stage	Early Production Stage	Speech Emergence Stage	Intermediate Fluency Stage	Near Proficient
	Beginning		Intermediate	Near Proficient	
L_2 Language Characteristics	- minimal comprehension - no verbal production - communicate with action and gestures	-limited comprehension -one/two word responses -short phrases	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech -engage in conversation and produce connected narrative	-near proficient orally *very good comprehension *vocabulary approaches that of native speaker of same age *few, if any, errors in speech *understands and produces complex sentences expected of native speaker of same age -lack of experiences with written language -scores at or above 40th percentile on state approved achievement test
Performance Indicators	listen draw point select move choose mime act/act out match circle	name list label categorize group tell/say respond answer (1-2 words)	recall summarize retell describe define role-play explain restate compare contrast	analyze evaluate create justify defend support debate	All performance indicators
Instructional Focus	• Develop listening comprehension • Build receptive vocabulary	• Expand receptive vocabulary • Design activities that motivate students to produce vocabulary and structures they already understand	• Continue to expand receptive vocabulary • Design activities to develop higher level of language use • Introduce language experience activities	• Continue to expand receptive vocabulary • Design activities to develop higher levels of language use in the content areas • Incorporate reading and writing activities into lessons	• Continue to expand receptive vocabulary • Design activities to develop higher levels of language use • Incorporate reading and writing into lessons • Introduce contrastive analysis • Introduce easy-to-read books • Provide teacher and peer support (e.g. cooperative learning)

Grade 5

Integrated ENGLISH LANGUAGE DEVELOPMENT(ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limited comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, move, choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend, support, debate, examine, complete, describe (detail)
	PRE-PRODUCTION STAGE	EARLY PRODUCTION STAGE	SPEECH EMERGENCE STAGE	INTERMEDIATE FLUENCY STAGE
MATHEMATICS	Raises visual of numeral (0-20) upon hearing name Picks up objects and sorts into two groups: heavy/light, clean/dirty Sorts three identical pictures into 3 categories: short, shorter, shortest and tall, taller, tallest Matches appropriate numeral to picture of corresponding number of objects	Sequences cards of numerals 0-20 from smaller to larger Names appropriate numeral 0-10 upon seeing visual reflecting corresponding number of objects Responds with "more than" or "less than" when shown two different numerals 0-10 Labels groups of objects with corresponding numeral Names geometric figures	Counts to 15 Associates appropriate numeral to corresponding picture of 10-15 objects Identifies objects in the room with specific geometric shapes Compares sets of two objects/persons for greater height and/or weight Retells steps of addition/subtraction problem leading to solution Demonstrates and explains concept of "greater than" and "less than" with two sets of objects	Counts to 20 Associates appropriate numeral to corresponding picture of 15-20 objects Makes a complete statement about the geometric shapes of various objects in the room/home/school Examines three objects/persons and describes them in relation to one another Reads and writes 0-20 Creates and solves oral word problems using objects
SOCIAL STUDIES	Points to self upon hearing name called Shows picture of appropriate family member upon hearing name Shows age and ages to be with object counters Raises hand when birthdate is called Listens to simple directions and places object in appropriate position Acts out one responsibility of family member names Circles appropriate day on calendar upon hearing day of week	Labels family positions in a personal photograph of own immediate family Tells birth date Tells current age and age to be next birthday Responds with location word only when asked position of object to another Names family member when household responsibility is given Says days of week in order	Describes self in relation to either family members Explains past and future events in own life Defines specific roles of various family members Contrasts responsibilities of different family members Answers with appropriate day when an event is described Recalls at least three rules of the classroom Explains location of objects in room	Describes other family members in relation to self Completes an open-ended statement about various family members Creates and describes an imaginary birthday Examines ways to obtain food and clothing Supports need for protection against heat and cold Compares and contrasts self to a favorite family member Justifies need for two classroom rules Explains need for safety practices
SCIENCE	Points to the appropriate sense organ when cued orally as to its sense and function Dramatizes an appropriate reaction to a visual depicting a weather situation Moves or touches the appropriate body part when its name is stated Chooses the appropriate sign "living" or "non-living" when shown various objects or plants/animals/people Follows commands correctly involving body parts	Names appropriate sense when cued verbally of function States body part(s) responsible for each sense Lists all body parts Indicates with one/two words a weather term reflected by various visuals Labels various pictures with terms dirty/clean or safe/dangerous	Describes weather of the day in a simple sentence States one function of each of the five senses Divides thing into categories of "living" and "non-living" while naming it Explains how many of each body part an individual has Tells about objects in the classroom/school/ home that represent terms clean or dirty or safe or dangerous	Compares and contrasts today's and yesterdays weather States one function performed by each of the body parts Describes what the loss of a sense might be like Develops criteria for identifying "living" and "non-living" Examines the advantages of good personal hygiene

Grade 1

Integrated ENGLISH LANGUAGE DEVELOPMENT (ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limited comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, move, choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend, support, debate, examine, complete, describe (detail)
	PRE-PRODUCTION STAGE	EARLY PRODUCTION STAGE	SPEECH EMERGENCE STAGE	INTERMEDIATE FLUENCY STAGE
MATHEMATICS	Holds up set when hearing numerals 0-20 Points to numbers 1-100 according to verbal cue. Selects largest or smallest number to 100 when asked. Holds up correct math symbol when asked. Shows penny, nickel, dime or quarter when asked. Picks out half of a group or region when asked. Tells place value of a 2-digit number.	States number from 0-20 when seeing set or numeral. Counts to 100. Writes numeral 1-100 upon hearing verbal cue. Names math symbols. Identifies penny, nickel, dime and quarter by stating correct name. Tells value of coins: penny, nickel, dime and quarter. Counts to five with ordinal numbers.	States steps to solving addition and subtraction problems with answer Solves oral word problems using addition and subtraction. Tells time to the hour and half hour. Measures with non-standard units. Counts coins to 50 cents. Uses ordinal numbers though 5 in a simple sentence.	Creates and solves written word problems using addition and subtraction. Uses names of standard and non-standard measurement in a sentence. Recognizes ordinal numbers through 5 in their written form. Uses names of money denominations in a complete sentence.
SOCIAL STUDIES	Points to appropriate picture of family member when given verbal cue. Holds up pictures of school personnel when requested. Points to appropriate picture when given verbal cues of national/cultural symbols and holidays. Points north, south, east, west on a map according to verbal cue. Follows simple directions utilizing relative location (spatial) words. Acts out roles of family members and school personnel.	Names an appropriate family member when a visual is shown and corresponding role is stated. Says name of school personnel upon seeing visual and hearing role described. Answers "yes" or "no" if statements heard promote cooperation between friends and classmates. Names national/cultural symbols and holidays when shown appropriate pictures. States "north/south/east/west" while observing such direction.	Follows two consecutive directions. Tells the role of each family member in a simple sentence. States one job responsibility of each school employee. Lists basic needs of a family. Describes three ways to promote cooperation at word and play. Responds to questions about safety rules. Gives other directions using relative direction words.	Describes in detail an individual responsible for making home, school, and community better. Explains diversity among individuals, families, and cultures. Identifies ways that the physical environment affects how people live. Gives oral directions to others using north/south/east/west. Tells about safety practices personally used in getting to and from school.
SCIENCE	Raises hand when hearing name of living things. Points to picture of season according to verbal cue. Draws sky, sun, cloud, and water when requested. Points to weather pictures according to verbal cue. Touches sense organ when cued orally for name and function. Points to plant part as directed. Holds up magnet, mirror, magnifying lens when asked.	Responds "living/non-living" when seeing or hearing cue. names the four seasons. Lists what living things require to sustain life. Indicates "day/night" when shown various pictures. Answers simple questions about weather with one/two word responses. Names plant parts. Lists animals when cued with a visual reflective of a certain habitat.	Describes each of the seasons in a simple sentence. Observes and retells steps of a simple experiment. Tells the functions of the various sense organs. Identifies various animals with a brief description of each. Defines what each part of a plant contributes to its life. Describes a proper diet and a poor diet.	Justifies individual daily intake of food as a "proper diet". States the relationship between proper diet and good health. Explains how to care for a pet. Compares and contrasts the seasons. Defines opposite characteristics as rough/smooth and gives examples.

Appendix F

Grade 2

Integrated ENGLISH LANGUAGE DEVELOPMENT (ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limited comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, move, choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend, support, debate, examine, complete, describe (detail)
	PRE-PRODUCTION STAGE	EARLY PRODUCTION STAGE	SPEECH EMERGENCE STAGE	INTERMEDIATE FLUENCY STAGE
MATHEMATICS	Raises numeral card 1-999 upon hearing verbal cue. Shows one color card for odd numbers and another color card for even numbers. Nods "yes" to indicate "correct" when hearing answers to addition and subtractions facts to 19. Points to half, thirds, or fourths of shaded diagrams upon hearing verbal cue. Selects penny, nickel, dime, quarter or half dollar when asked. Groups objects by 2's, 3's, 4's, 5's, 10's.	States "smaller/larger/equal to" when asked to compare two numbers. Says "odd" or "even" when seeing or hearing numbers to 999. Counts by ordinal numbers through 10. Names coins and states their value. Gives the answer to addition and subtraction facts to 19. Identifies by name these symbols: +, -, =, >, <. names shaded diagrams by half, thirds, and fourths. Counts by 2's, 3's, 4's, 5's, and 10's.	Read equations indicating "larger than/smaller than/equal to". Verbalizes addition and subtraction facts with answers up to 19. Tells time to the quarter hour. Measures using standard units while naming unit of measurement. Counts money while making change. Uses a math concept related to the calendar in a simple sentence. States a math concept in a sentence as it relates to real life experiences. Says a complete date, including month, day, and year.	Counts from 1-999. Explains concept of regrouping in own words. Uses math concepts to explain personal life experiences. Constructs and interprets simple graphs and tables. Expresses steps of an experience by using ordinal numbers (first, second).
SOCIAL STUDIES	Holds up picture of appropriate community helper according to verbal cue. Touches picture of facility or service available upon hearing name of place in community. Points to own country, state, and city on map when asked. Draws map symbols when given oral direction. Demonstrates appropriate response to simple classroom rules.	Names community members on picture cards. Gives name of community facility when a service is stated. Responds with one/two words to simple questions about community people, facilities, and services. Replies with "a need" or "a want" upon hearing statement indicative of either. States name of own country, state and city.	Gives an example of a rule and a law in a simple sentence. Names community members and the respective service each provides. Identifies groups in which people have specific roles. Uses map symbols and keys. Answers questions about services and facilities provided in the community.	Explains why society needs rules. Defines criteria for distinguishing between "a need" and "a want". Describes how the passage of time effects people and living things. Tells about one individual who has made a contribution to the community. Creates and tells about a new facility or service for the community.
SCIENCE	Shows appropriate picture of plant or animal when requested. Raises hand when hearing the name of living things. Points to pictures of energy sources when hearing verbal cue. Holds up appropriate food group card when hearing name of food. Points to appropriate picture of solid, liquid, or gas according to verbal cue. Selects correct environment picture card when hearing name of various animals.	Names plant or animal on picture card. Diagrams plant stages with simple directions. Sequences animal life steps with simple directions. States "solid/liquid/gas" upon hearing or seeing cue. names energy sources when shown pictures. Identifies orally appropriate food group when shown picture of various foods.	States the importance of water to living things in simple sentences. Answers questions about animals and their environment in complete sentences. Verbalizes the sequence of stages of plant and animal life. Makes a chart with food groups listing various foods in each category.	Classifies animals into environmental categories while describing each habitat. Compares prehistoric life to life today. Predicts outcomes in cause/effect situations. Describes the three states of mater. States the stages of plant and animal life using ordinal numbers throughout the sequence.

Grade 3

Integrated ENGLISH LANGUAGE DEVELOPMENT (ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limited comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences --complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, move, choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend, support, debate, examine, complete, describe (detail)
	PRE-PRODUCTION	EARLY PRODUCTION STAGE	SPEECH EMERGENCE STAGE	INTERMEDIATE FLUENCY STAGE
MATHEMATICS	Selects appropriate numeral (1 to 9,999) according to verbal cue Points to appropriate math symbol upon hearing name of operation Points to appropriate coin/paper money upon hearing name Selects correct numerical answer to multiplication and division facts through 45 with verbal cuing. Choose correct answer for multiples of 10 and 100 with verbal cuing Place hands of clock on correct numbers for time stated Recognizes appropriate devices for measuring	Names numerals (1 to 9,999) States math operation (+,-,x,/) upon seeing a completed problem Names different coins and paper money denominations Verbalizes answer to multiplication and division facts through 45 States answer to problems involving multiples of 10 and 100 Tells time to the nearest hour Labels and categorizes various devices for measuring with either traditional or metric systems	Uses numerals orally (1 - 9,999) in context of a simple sentence Explains math operation using name of math symbol. Counts change correctly States answer to multiplication and division facts through 45 Verbalizes multiples by 10 and 100 Demonstrates commutative property Tells time to five minute intervals Compares and contrasts orally appropriate devices used to measure with traditional and metric systems	Reads and writes numeral 1-9,999 Explains math operation involved for solution to 3-digit numerals Counts change through addition and subtraction Verbalizes multiplicand, multiplier and product or multiples through 45 and by 10 and 100 Defines commutative property Describes in detail (5 min. intervals) all personal events of previous day Measures using traditional and metric systems with appropriate devices
SOCIAL STUDIES	Points to self upon hearing own cultural/ethnic group named Draws land formation upon hearing geographic name Shows labels "density/sparsity" of visual reflecting an area of the city Observes pictures reflecting four different seasons Points to forms of transportation upon hearing name Follows map upon hearing simple spatial directions Matches community workers with appropriate place of employment	Responds with name of land formation upon seeing visual cue Says "density" or "sparsity" upon seeing visual of a specific area of the city Names the appropriate seasons when shown visual of weather Identifies name of community worker upon seeing visual cue Lists modes of transportation available in the region Responds "right" or "wrong" when listening to directions from own home to school	Gives name of friend and cultural ethnic group in a complete sentence Describes location of home in relation to a specific land formation Paraphrases meaning of population density/sparsity Gives directions from home to a specific place in the city. Describes the job of one community helper Recalls all forms of transportation describing a preference Explains the meaning of pollution Compares specific sites of the City in relation to periods of time	Analyzes the advantages of the diverse ethnic population of the City/Region Evaluates the contributions of one cultural group of the City/Region Explains why land formations affect population density Completes a sentence describing the function of a local government unit when hearing its name Examines a map of the region describing key community places Debates the advantages/disadvantages of major industry in the City/Region Tells about favorite season of year
SCIENCE	Points to appropriate plant or animal upon hearing name Matches certain plant and animal population to their corresponding habitats Circles objects that reflect a solid state, underlines pictures that are liquids, and puts an x over pictures of gases Points to mountains, rocks, and soil in an appropriate picture Selects food when called by name.	Labels appropriate animals as "extinct/endangered" upon hearing names Names plants and/or animals that live in the natural habitat reflected by a picture. Categories numerous pictures as either solid, liquid, or gas. Groups food according to their major most common nutrient. Calls mountains, rocks, and soil by correct names in an appropriate picture.	Describes why a specific plant or animal is an endangered species. Paraphrases the natural habitat of two animals. States "photosynthesis" in three sentences. Recalls events estimating specific numbers and time Defines a solid, a liquid, and a gas Explains two simple machines Compares and contrasts two animals Describes favorite foods based on nutritive value	Defends extinction of an endangered species Defines an animal's habitat including climate, land formation, surroundings, plant life, and other animals Describes in detail plant needs for proper growth Justifies conservation practices Analyzes criteria for a solid, a liquid, and a gas Examines contributions of three major scientists

Grade 4

Integrated ENGLISH LANGUAGE DEVELOPMENT (ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limited comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, move, choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend, support, debate, examine, complete, describe (detail)
	PREPRODUCTION STAGE	EARLY PRODUCTION STAGE	SPEECH EMERGENCE STAGE	INTERMEDIATE FLUENCY STAGE
MATHEMATICS	Selects appropriate numeral (1 - 99,999) according to verbal cue. Points to math symbol upon hearing name of a given operation (+,-,x,/). Shows correct answer with verbal cuing to multiplication and division facts through 10 x 10. Places hands of clock on correct numbers for each time stated. Points to appropriate coin with each verbal cue. Circles appropriate Roman Numeral upon hearing numerical name. Points to geometric component with verbal cue.	Names numerals (1-99,999). States correct math operation upon seeing a completed problem. Tells time upon seeing clock. Verbalizes correct answers to multiplication facts up to 100. Names all coins to make change. Gives appropriate names of geometric components. Restates Roman Numeral in Arabic terms. Labels a line, line segment, point, ray, and angle. Groups separately fractions, decimals, and whole numbers.	Uses numeral (1 - 99,999) in a simple sentence. Explains math operation using name of math symbol in explanation. Describes events with reference to specific time (hour and minutes). makes change with verbal explanation. Tells how solution to a fraction was achieved. Compares equivalent fractions. Identifies the quotient, divisor, dividend and remainder of completed division problems.	Reads and writes (1 - 99,999) in simple sentences. Justifies answer to math problem naming math symbol. Reads time to nearest minute. Examines incorrect change received. Reads thermometer to nearest 10 degrees. Describes math operation for solving problems with decimals through hundredths.
SOCIAL STUDIES	Matches appropriate cultural group of region with verbal cue. Points to appropriate early inhabitant, explorer, or missionary. Chooses appropriate geographical term for visual cue. Points to appropriate part of government upon hearing function. Identifies appropriate visual of a natural resource with verbal cue. Acts out a safety measure in response to hazard cue.	Names diverse cultural groups of region. Categorizes individuals into 3 groups: early inhabitants, explorers, or missionaries. Labels visual with appropriate geographical form. Labels appropriate part of government upon hearing function. Responds with one/two words to a hazard (help, run, stop).	Describes each cultural group in region with two simple sentences. Explains the life experiences of the early habitant, the explorer, and the missionary. Defines the major geographical terms/features. Tells the function of each part of state government. Role-plays (verbal) reactions to safe/hazardous situations.	Examines in detail the contributions of one cultural group in region. Tells the autobiography of one early inhabitant, explorer, or missionary. Creates an undiscovered land mass and describes it with geographical terms. Describes the interrelationship of the parts of the state government. Analyzes appropriate/inappropriate reactions to safe/unsafe situations.
SCIENCE	Follows simple experiment directions. Identifies simple machines with verbal cue. Places the planets of the solar system in correct location to one another. Matches weather terms to appropriate visual cue. Draws appropriate body organ for each verbal cue. Points to appropriate material to construct an electrical current (battery, wire, bulb).	Names the planets of the solar system. Labels weather terms for each visual cue. Responds with correct name for various body organs. Lists all materials to construct an electrical current. Names simple machines.	Describes the solar system using appropriate planet names. Defines weather terms as air pressure, heat, cold, etc. Recalls component parts of major body organs. Explains how to construct an electrical current. Recalls major effects of drug abuse.	Describes why one has not yet gone to a particular planet. Justifies today's weather using weather terms. Analyzes functions of major body organs. Evaluates another's construction of an electrical current. Explains interrelationship between plants and animals.

Grade 5

Integrated ENGLISH LANGUAGE DEVELOPMENT (ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limit comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, move choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend, support, debate, examine, complete, describe (detail)
	PRE-PRODUCTION STAGE	**EARLY PRODUCTION STAGE**	**SPEECH EMERGENCE STAGE**	**INTERMEDIATE FLUENCY STAGE**
MATHEMATICS	Holds up appropriate numeral card (1 - 100,000,000) according to cue Points to place value (1 - 100,000,000) according to cue Shows understanding of terms "decimal/fraction/whole number" by pointing according to verbal cue Chooses appropriate fractional part of diagram when given oral cue Draws parallel and perpendicular lines according to oral cue Identifies parts of fraction (nominator and denominator and like/unlike denominators)	Names place value of ones, tens, hundreds, thousands Responds orally with "fraction", "whole number", or "decimal" when shown flashcard Responds orally with "greatest common factor" or "least common denominator" when given a list of numbers with one number shaded for distinction Names fractional parts when shown shaded diagrams Lists metric units of measurement from smallest to largest Labels diameter and radius	Restates numbers in expended notation Renames improper fractions and mixed numerals Tells weight of object measured on scale Measures and reports diameters and radius of flat, round objects Compares types of triangles Reads selected numerals (1 - 100,000,000) aloud	Describes the calculation of averages Explains metric conversions Explains use of protractor while measuring angles Compares congruent, symmetric and similar polygons and segments States all steps to solve a 3-place multiplication problem
SOCIAL STUDIES	Matches cultural group card to correct name upon hearing verbal cue Sequences in order first four periods of our country's history Points to river, mountain, lake, ocean, and desert upon receiving verbal cue Matches picture with name of natural resource Matches capital with state Selects correct map type when named	Names periods of U.S. history in natural sequence from earliest to present times Gives name of geographic features when shown various ones on a map Lists major natural resources in the U.S. States capital when given state Names types of maps Lists major cultures of U.S. while matching a visual of a contribution or influence of each	Explains the parts and functions of national government Defines democracy and its principles in simple sentences Compares the use of different types of maps Defines "ecological balance" in a complete sentence compares and contrasts two distinct periods of U.S. history	Describes a favorite period of U.S. history Supports the need for ecological balance Analyzes the contributions and influences of U.S. diverse cultures Defends democracy against other forms of government Explains use of one specific map based on a purpose
SCIENCE	Points to picture of energy source when given verbal cue (solar, nuclear). Chooses simple machine indicated by verbal cue Selects appropriate picture of animal or plant when receiving verbal cue Selects correct food group when given food name orally Holds up visual of stage of life cycle of insect when hearing name	Categorizes plants and animals by classification group Names stages of photosynthesis when shown pictures Labels stages of water cycle Names planets in solar system Labels stages of social insect life cycle Names four food groups, giving an example of each	Explain steps of scientific problem-solving in simple sentences States the steps of scientific decision-making in complete simple sentences List steps of performing an experiment Defines energy, conservation and natural resource Explains the use of each simple machine	Compares science problem-solving and decision-making Describes the characteristics of the solar system's celestial bodies Explains the process of photosynthesis Defends the need for conservation of energy and national resources Creates a chart showing major developmental changes during puberty Creates a balanced meal using major food groups in the rational for choice

Grade 6

Integrated ENGLISH LANGUAGE DEVELOPMENT (ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limited comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, move, choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend support, debate, examine, complete, describe (detail)
	PRE-PRODUCTION STAGE	EARLY PRODUCTION STAGE	SPEECH EMERGENCE STAGE	INTERMEDIATE FLUENCY STAGE
MATHEMATICS	Chooses correct category of whole number, fraction or decimal upon receiving verbal cue. Selects improper fractions or mixed numbers upon verbal cuing. Picks out correct percent indicated verbally, when given a choice on paper. Points out angle, segment, triangle, square, and circle when signaled verbally.	Round out whole numbers and decimals orally. Names information needed to determine volume of a box. Names information needed to determine area of circle. Changes fractions and decimals to percent orally, with flashcard cue. Gives metric measurements of length, capacity and mass. Reads temperature in degrees Celsius.	Explains rounding out of whole numbers and decimals. Explains reduction of fractions. Defines triangle, circle, segment, box and angle. Defines mass capacity, length, and volume. Renames decimals as fractions and fractions as decimals.	Explains process of calculating area of circle. Describes process of calculating volume of a box. Compares ratios to determine equivalence. Explains process of converting decimals and fractions to percents. Predicts probable outcomes when given simple situations.
SOCIAL STUDIES	Selects picture of natural resources according to verbal cue. Points out geographical features on a map when cued orally. Points to graph, chart, table or timeline when cued verbally. Identifies reference materials when called for such as atlas, dictionary, encyclopedia.	Names natural resources. Identifies geographical features on a map. Shows movement of people on a map according to directions given. Labels stages of trade development. Names reference materials used in social studies class. Lists characteristics of culture.	Lists the factors that constitute a political system. States the factors that constitute an economic system. Defines colonization, settlement, and civilization in own words. Provides examples of technology, politics, and economy in short sentences.	Describes how climate, natural resources, and physical features affect settlement (culture). Compares the colonization of U.S. with another country. Compares the technology of yesterday with that of today. Explains the use of map key or legend. Interprets information on a graph, chart, table, or timeline. Creates a timeline of trade development stages. Uses special maps and globe projections to acquire information.
SCIENCE	Sequences picture card events of a demonstration after observation. Identifies chart, graph, table and diagram by pointing following verbal cue. Copies chart, graph, table and diagram. Matches picture of animal/plant to appropriate name with verbal cuing. Points to male and female reproduction system, when asked verbally, on an anatomical chart.	Selects appropriate instrument after hearing description of purpose. Tells mass of length of object in metric units. Follows directions to perform simple investigations. names non-typical plants and animals when shown picture. Names organ systems.	Sequences the events of a teacher demonstration. Interprets simple graphs, charts, tables and diagrams to answer basic questions. Lists scientific instruments and their uses. Categorizes plants and animals into typical and non-typical types.	Gives directions for a simple investigation to another student. Creates a graph or chart based on information given. Describes the functions of various organ systems. Differentiates between fact and opinion based on scientific method. Explains functions of male and female reproductive systems. Supports conclusions drawn with observations made.

BEST COPY AVAILABLE

Appendix K

Grade 7

Integrated ENGLISH LANGUAGE DEVELOPMENT (ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limited comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, mime, choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend, support, debate, examine, complete, describe (detail)
	PRE-PRODUCTION STAGE	EARLY PRODUCTION STAGE	SPEECH EMERGENCE STAGE	INTERMEDIATE FLUENCY STAGE
MATHEMATICS	Selects correct power expressed by exponent with verbal cuing. Points to a fraction, whole number, mixed number, or percent according to oral directions. Identifies a correct number when asked "prime" or "composite". Holds up appropriate figure when asked for congruent, similar or 3-D figures. Points to correct spot on graph when orally given coordinates.	Responds orally with "fraction", "decimal", "mixed number", or "percent" when shown a flashcard. Categorizes a list of numbers under "prime" and "composite" headings. Identifies congruent, similar and 3-D figures orally when shown figure. Tells the length of an object in metric units. States the surface area of a 3-D figure.	Lists the information needed to calculate surface area. Restates fractions as decimals, and decimals as fractions. Restates decimals as percents and percents as decimals. Predicts outcome of a given situation. Locates information requested on a graph.	Explains the conversion of fractions to decimals and decimals to fractions. Describes the process of changing percent to decimal and decimal to percent. Describes the process of finding what percent one number is of another. Explains how to find the opposite of an integer. Outlines steps for finding the mean. Defines "prime" and "composite".
SOCIAL STUDIES	Selects appropriate picture when hearing geographical feature. Points to picture of map, globe, chart, or graph according to verbal cue. Selects correct symbol (predetermined when hearing the terms "social" (people), "political" (document), "economic" (money), "intellectual" (head), and "cultural" (flags). Acts out according to verbal cue "drawing", "modeling", "writing", "discussing", "reporting", and "dramatizing".	Sequences pictures of geographical changes upon hearing simple description. Responds with "yes" or "no" to questions posed orally about simple chart. Locates area on globe or map when asked orally. States "discussing", "reporting", "dramatization", "drawing", "modeling", or "writing" after seeing demonstration. Draws a replica of world culture today and a prediction of tomorrow's world culture.	Names the major geographic epochs. Distinguishes verbally between "prehistoric" vs "historic". Names significant geographic features. Defines "conflict" and "controversy". Makes a chart of several great ideas and inventions that have transcended time and place, including origin, function, importance. Names men and women who have made significant contributions while stating their contributions. Creates a chart depicting major geographical epochs.	Explains the geographic transition from prehistoric to historic times. Relates geographical features to development of civilization. Compares civilization of whole world with civilization of non-western world. Reports on a conflict or controversy between cultures or nations. Takes notes on information regarding U.S. political system - listing major characteristics. Diagrams the effect that technological change has on cultural change. Formulates questions regarding political system of another country.
SCIENCE	Points to correct picture upon hearing "living" or "non-living". Points to major parts of microscope according to verbal cue. Sequences pictures for properly using the microscope. Holds up correct numeral (1-4) upon hearing corresponding part of experiment.	States "living" or "non-living" when shown pictures. Identifies major parts of microscope. Name four parts of an experiment. Diagrams human activities and their effect on the environment. Follows directions for using a microscope.	Describes the parts of an experiment hypothesis, procedure, results, conclusion. Lists ways the environment effects human activity. Gives directions for using a microscope. Tells volume of object in metric units.	Describes "cooperative group behavior" in terms of class experiments. Collects data and organizes it in column or graph form. Explains the effect that human activity has on the environment. Defends the statement "We must appreciate and respect the environment now, or someday our carelessness will take its toll".

Grade 8

Integrated ENGLISH LANGUAGE DEVELOPMENT (ESL) in the Content Areas

PROFICIENCY	LEVEL I	LEVEL II	LEVEL III	LEVEL IV
LEP PUPIL DESCRIPTORS	-minimal comprehension -no verbal production	-limited comprehension -one/two word response	-increased comprehension -simple sentences -some errors in speech	-very good comprehension -more complex sentences -complex errors in speech
LEP PUPIL PERFORMANCE INDICATORS	listen, draw, point, select, move, choose, mime, act/act out, match, circle	name, list, label, categorize, group, tell/say, respond, answer (with one/two words)	recall, summarize, retell, describe, define, role-play, explain, restate, compare, contrast	analyze, evaluate, create, justify, defend, support, debate, examine, complete, describe (detail)
	PRE-PRODUCTION STAGE	EARLY PRODUCTION STAGE	SPEECH EMERGENCE STAGE	INTERMEDIATE FLUENCY STAGE
MATHEMATICS	Points to correct numerical or algebraic expressions according to verbal cue Selects circle, square, triangle or rectangle when hearing its name Points to parallel or perpendicular line upon hearing oral cue Locates spot on graph when hearing coordinates	Selects correct number written in science notation, upon hearing it Responds with "yes" or "no" when given oral estimates of area, volume, mass and temperature Estimates orally the area, volume, mass or temperature Names geometric shapes - triangle, square, rectangle, circle Labels parallel and perpendicular lines. Tells number of permutations or combinations Identifies mean, median, and mode	Reads numbers written in scientific notation Lists steps for prime factorization Provides information necessary to calculate area, volume, mass and temperature Defines "circumference" in own words. Defines "parallelogram" in own words. Lists steps for finding circumference of a circle and area of a triangle and parallelogram Locates information requested in graphs - mean, median, and mode	Explains the simplification of a numerical and an algebraic expression Interprets graphs to provide requested information in written form Contrasts purpose of "mean", "median", and "mode" Describes the process for finding a square root
SOCIAL STUDIES	Selects correct picture when given geographical feature orally Points to appropriate picture when hearing name of cultural group in U.S. Responds with "yes" or "no" when shown the picture of a contributor to U.S. cultural development and a contribution picture, indicating match or no match	Names men and women who have contributed to the U.S. upon hearing a specific contribution Sequences pictures of cultural changes from Colonial to present day America while hearing a simple description Shades in a map of the U.S. showing growth while hearing land acquisition story Responds "yes" or "no" to statements relating to contributions of slave society to U.S. culture	Creates a chart showing effect of geographical features on settlement patterns Expresses reasons for exploration and colonization, as well as problems Lists major features of the Constitution and values and beliefs the represent Diagrams relationship between invention, technological change and social problems	Compares and contrasts Colonial values with contemporary values compares and contrasts two major periods in U.S. immigration Writes a diary entry from the viewpoint of an immigrant, explaining problems he/she has faced Describes the U.S. economic structure and its role in world economics Writes an essay on "The Role of the U.S. in the World Today: From a Global Perspective" Predicts what American culture will be like in the future
SCIENCE	Performs an experiment with picture and simple verbal directions. Shows appropriate piece of equipment when cued verbally Points to cardiovascular, reproductive and respiratory systems on an anatomical chart, according to verbal cue	Performs an experiment with simple verbal directions Follows simple oral directions for using science equipment Pairs disease with appropriate system indicating place of occurrence. Names potential disease when hearing the name of an organ	Lists methods and alternatives to conserve and preserve earth's physical properties Performs experiment with written directions Chooses manner in which data should be presented when given experiment situations States the steps followed in an experiment Charts systems and common problems/diseases	Designs the experiment to test a hypothesis Creates an information/instruction sheet for the use of science equipment. Expresses an opinion on a contemporary issue, supported by scientific knowledge Justifies choice of data presentation tool

Appendix M

Name of Student: ____________________

INDIVIDUAL ENGLISH LANGUAGE ASSISTANCE NEEDS FORM

Criteria	PK	K	1	2	3	4	5	6	7	8	9	10	11	12
	91-92	92-93	93-94	94-95	95-96	96-97	97-98	98-99	99-00	00-01	01-02	02-03	03-04	04-05
Home Language Survey														
OLPT														
Oral Lang. Interview														
Observation Data														
SES														
Schooling Experience														
Observation Data														
Achievement Test														
Cloze Test														
Six Wks. Grades														
Observation Data														
TOTAL Points														
Classification														
Parental Approval/Denial														

Appendix N

Campus English Language Assistance Needs Form

Campus: ______________

Grade Level: ______________

Student	Lang	Oral Language Criteria				Social Criteria			Academic Criteria				Total	Category	Parent A/D*
		Crit.1	Crit.2	Crit.3	Crit.4	Crit.1	Crit.2	Crit.3	Crit. 1	Crit.2	Crit.3	Crit.4			

* Parental Approval (A)
Parental Denial (D)

Appendix O

Likert Rating Scale for Determining English Language Assistance Needs

Oral Language Proficiency Data

Home Language Survey

1 – Only Native Language Spoken
2 – Mostly Native Language Spoken
3 – Native and English Languages Spoken
4 – Mostly English Spoken
5 – Only English Spoken

Oral Language Proficiency Instrument

1 – Non-English Speaker
2 – Extremely Limited English Proficiency
3 – Limited English Proficiency
4 – Near Native-Like English Proficiency
5 – Fluent English, Native-Like Proficiency

Interview Instrument

1 – 80-100% Native Language Responses
2 – 50- 79% Native Language Responses
3 – < 50% in either Language
4 – 50- 79% English Language Responses
5 – 80-100% English Language Responses

Observation Data

1 – Pre-Production Stage
2 – Early Production Stage
3 – Speech Emergence Stage
4 – Intermediate Stage
5 – Fluent Stage

Social Data

Socio-Economic Status

1 – < $5,000
2 – $5,000 - 10,000
3 – 10,000 - 25,000
4 – 25,000 - 35,000
5 – 35,000 - 45,000

Schooling Experience

1 – No Previous Schooling or All English Program Only
2 – Interrupted Schooling/Some ESL Instruction
3 – Schooling in Other Countries
4 – ESL program only since entering U.S. school system
5 – Bilingual education program only since entering U.S. school system

Observation Data (Home, with friends)

1 – Uses native language ONLY in all settings
2 – Relies on native language in all settings
3 – Uses the native language sparingly in all settings
4 – Uses the English language with friends only
5 – Uses the English language mostly in all settings

Academic Data

Standardized Achievement Data

1 - < 20 %ile
2 - 20-29 %ile
3 - 30-40 %ile
4 - 41-59 %ile
5 - 60-80 %ile

Cloze Test

1 – Raw Score of 0 - 20
2 – Raw Score of 21 - 30
3 – Raw Score of 31 - 40
4 – Raw Score of 41 - 49
5 – Raw Score of 50

Six Weeks Grades

1 – < = 59
2 – 60's
3 – 70's
4 – 80's
5 – 90's

Observation Data by Grade Level and Subject Area

1 – Points, identifies
2 – Names, lists
3 – Describes, tell (simply)
4 – Compares, describes (more complex)
5 – Analyzes, synthesizes

BEST COPY AVAILABLE

Notes

Acknowledgements to the following individuals for information contained in the Appendices and included in this paper.

Appendices A-B	Dr. Eleanor Thonis
Appendices C-L	Dr. Betty Mace-Matluck and the Southwest Educational Development Laboratory Multifunctional Resource Center; Austin, Texas

References

Bernstein, D. K. (1989). Assessing children with limited English proficiency: Current perspectives. Topics in Language Disorders, 9, No. 3, 1989, pp. 15-20.

Canales, J. (1989). Assessment of language proficiency: Informing policy and practice. Focus. Austin, TX: Southwest Educational Development Laboratory.

Christner, C. & Baenen, N.R. (1988, April). Longitudinal effectiveness of pre-kindergarten programs: Different approaches. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA,.

Cummins, J. (1980) The entry and exit fallacy in bilingual education. NABE Journal, 3, 25-29.

Erickson, J. G. & Omark, D. R. (1981) Communication assessment of the bilingual bicultural child. Baltimore: University Park Press.

Fishman, J.A. (1972). The sociology of language. In P.P. Giglioli (Ed.), Language and social context. Great Britain: Penguin Books.

Hernandez-Chavez, E., Burt, M., & Dulay, H. (1978). Language dominance and proficiency testing: Some general considerations. NABE Journal, 3, (1).

Louisiana State Department of Education. (1991). Results of 1990-1991 survey of limited English proficient students.

Mace-Matluck, B. (1988) Consequences of exiting students too early from special language assistance programs. Austin, TX: Southwest Educational Development Laboratory.

National Advisory and Coordinating Council on Bilingual Education. (1986). New directions in late '80s pursued (Tenth Annual Report, 1985-86). Washington DC: Author.

New Mexico State Department of Education. (1991). Annual report for limited English proficiency population, 1989-1990.

New York State Education Department. (1990). Guidelines for Programs Under Part 154 of Commissioner's Regulations for Pupils with Limited English Proficiency. Albany, NY: Author.

Oller, J. W. (1973). Discrete-point tests versus tests of integrative skills. In Focus on the learner: Pragmatic perspectives for the language teacher. Rowley, MA: Newbury House.

Olson, R. E. (1989). Report of the limited English (LEP) student enrollment survey.

Olson, R. E. (1991). Report of the 1991 limited English proficient (LEP) student enrollment survey.

Personal correspondence Re: Information on LEP Student Eligibility and Language Assessment Practices. Georgia Department of Education, Atlanta, Georgia, 1991.

Personal correspondence Re: Information on LEP Student Eligibility and Language Assessment Practices. The Commonwealth of Massachusetts, Department of Education, Quincy, Massachusetts, 1991.

Posner, B. et al. Dual Literacy Programs for LEP Students (Report). New York: New York City Board of Education, Division of High Schools.

Rasp, A., Jr., & Ensign, G. B., Jr. (1991). Washington State Transitional Bilingual Instruction Program (Annual Evaluation Report, 1989-90).

Report of the Committee for the Evaluation of Language Assessment. Austin, Texas: Texas Education Agency, 1977, 1979, 1983, 1985, 1988.

Rivera, C. (Ed.). (1984). Placement procedures in bilingual education and policy issues. Clevedon, England: Multilingual Matters.

Rivera, C. & Simich, C. (1981). Issues in the Assessment of Language Proficiency of Language Minority Students. NABE Journal, 6, (1).

Smith, M. E. & Heflin, J. F. (1988, April). An analysis of ESL/bilingual education policy in Oregon school districts. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Texas State Education Agency. (1977, 1979, 1983, 1985, 1988). Report of the committee for the evaluation of language assessment. Austin: Author.

Texas State Education Agency. (1990). Data and information on the educational condition of limited English proficient students in Texas for the 1989-90 school year. Austin: Author.

Texas State Educational Agency. (1991). State board rules of education. Austin: Author.

Thonis, E. (1980, July). Language assessment: A malignant good. Speech delivered in Honolulu, Hawaii.

Ulibarri, D. M., Spencer, M.L. (1981). Language proficiency tests and their relationship to school ratings as predictors of academic achievement. NABE Journal, 3.

Utah State Office of Education. Survey form for data on limited English proficient students in public and private schools of Utah.

Wentling, R. M. (1985). Identification, assessment, and placement procedures used for limited English proficiency (LEP) students in vocational education programs. Journal of Vocational Education Research, 10, (1), 41-57.

Response to JoAnn Canales's Presentation

Julia Lara
Council of Chief State School Officers,
Washington, DC

The comments outlined below draw extensively from the work of Ed De Avila, and from a report recently completed by the Council of Chief State School Officers (CCSSO) titled, "Recommendations for Improving the Assessment and Monitoring of Students with Limited English Proficiency".

The author of this paper should be commended for bridging the gap between our knowledge of socio-linguistic theory of language learning and the application of the principles of this theory to the assessment of limited English proficient students. There has been for a number of years agreement within the field concerning the need to encourage the use of integrative approaches to language assessment (observations, interviews, dictation, etcetera. However, as noted in the paper, these approaches can be costly and time consuming and consequently districts have been reticent to use these approaches extensively. Another key barrier preventing the use of these approaches has been the absence of an operational definition of a limited English proficiency student and of a fully English proficient student (see CCSSO document for conceptual definition of LEP and FEP)[1]. The methods of assessment outlined in the paper; the rating scales; and the social and academic data elements suggested are important elements of a comprehensive data collection system on LEP students.

However, there are a number of areas that need clarification and perhaps elaboration in this paper. The following comments discuss each of this areas of concern.

The discussion of state assessment and data collection practices is limited given the limited number of survey responses obtained by the author. A more extensive discussion of state assessment and data collection practices is contained in a publication by the CCSSO titled, "Summary of State Practices Concerning the Assessment of Data Collection about Limited English Proficient Students." This report lists on a state by state basis, the pre-screening, classification, placement, and exiting procedures and instruments used in each state, and types of instrument. The report also identifies data elements collected at the state level on LEP performance and academic status.

Differentiation needs to be made between procedures used for classification of language proficiency status from those to be used for placement. It appears that the author suggests that integrative approaches used be used for purposes of classification along with the traditional oral language proficiency and achievement tests. While in the ideal situation this would be the best course of action to follow, we cannot loose sight of the realities (limitations in funding, personnel) at the local level, and the importance of identifying students within a reasonable period of time. In states with large numbers of LEP students, LEAs are advised to screen, classify, and place LEP students in language assistance programs within 30 days of enrollment. Districts in these states must use methods that are simple, effective, quick, and efficient. I am not convinced that, for purposes of classification, local practitioners can use all three assessment methods suggested (oral language proficiency tests, achievement tests and integrative tests) within 30 days or less. In spite of the limitations inherent in the language proficiency tests (do not measure all four language areas) districts may need to rely heavily on these instruments' use for purposes of classification[2]. However, it makes sense to use integrative approaches in borderline cases when student's score on the language proficiency tests are close to the cutoff point.

For purposes of placement, monitoring language development and mainstreaming LEP students into the English-Only classroom, it is essential that the communicative based approaches outlined in the model be used by classroom teachers on a consistent basis. These assessments are particularly important prior to decision-making points along the LEP student educational continuum. A review of state practices shows that, for placement purposes, no state requires the use of observations, although 33 states do recommend that these methods be used. In terms of interview methods in five states they are required, while in 23 states they are recommended. Unless these procedures are required by the state, it is difficult to sort out when and how districts use integrative approaches. It appears that in many instances, LEAs opt for the least expensive option. Thus, at the national level, we do not have a clear picture of local practice regarding use of various assessment instruments. However, we do know that LEAs with resources are more likely than others to use a variety of assessment methods for purposes of placement and exiting.

In terms of reclassification, there is no doubt that, at the classroom level, teachers need to have information about what students can and cannot do relative to the linguistic demands of the mainstream classroom. Without this normative information, placement decisions are likely to be made in isolation of the classroom context and may result in premature exiting of LEP students from the language support programs. Integrative methods are certainly the most valid mechanisms for providing information to teachers about student linguistic performance.

More needs to be said about issues of reliability. Some concerns have been raised about the extent to which rating scales can be applied systematically across various contexts. Ed DeAvila has noted in his writings that teacher rating is problematic because they are highly dependent on the teacher's language background, the teacher's familiarity with the child, and the teacher's knowledge of language development (Ed DeAvila, 1990). I am not certain that these concerns have been addressed by the model. Assessment experts may need to look more closely at this issue relative to the recommendations outlined in this paper.

The use of socio economic data for purposes of identification and placement can be misused. The author asserts that this information can be used as an indicator of "oral/aural stimulation received in the home" and subsequently suggests family income as the measure of SES. The relationship between lack of stimulation in the home and development of linguistic skills in the LEP students' first of second language needs further exploration. There is no direct relationship between poverty status and inability to learn a second language as there is between poverty and academic achievement (broadly defined). To imply that there might be a positive relationship between the two is to minimize the role of both the developing linguistic and literacy skills of LEP students independent of socio economic background. The author needs to strengthen the case for the use of SES as an important element of the profile and show how it bears on language learning.

In terms of data collection, the data elements contained in the ELAN profile will be useful in terms of classroom level instructional needs. However, for decision making at the state and local level, the data set needs to be more comprehensive. Administrators and decision makers need information that can be used for program evaluation/development purposes such as referrals to special education, placement in categorical programs, dropout rates, attendance, retention in grade and much more.

Finally, while this paper identified the key assessment methods essential for student identification, it did not outline how these various assessment methods would relate to each other and at what point in the educational experiences of the LEP student. Nonetheless, the ELAN Profile chart is a promising mechanism for decision making at the local level. With additional development it should be very useful to practitioners and to officials in state education agencies.

Notes

[1] The CCSSO publication cited above, "Recommendations for Improving....Limited English Proficiency," contains a definition for a limited English proficient student and for a fully English proficient student.

[2] Along with the information obtained in the screening devise, Home Language Survey.

Response to JoAnn Canales' Presentation

Robert Rueda
University of Southern California

The paper by Dr. JoAnn Canales on innovative practices in the identification of LEP students set out to accomplish three distinct goals. One was to provide information on current identification practices by state departments of education, including measures which they suggest or propose. A second goal was to present a way to systematically identify LEP students through the use of multiple alternative criteria. Finally, the last goal was to outline a paradigm that would permit state departments to collect consistent data for students in need of English language assistance. In preparing my comments, I have followed the order of the main points made in the paper, and therefore I will present those comments in that sequence. At the end of the commentary, I will present a set of suggestions for possible future drafts of the paper.

Current Practices

In order to provide data on current practices around the country, Dr. Canales conducted a survey in which seventeen states were contacted. Responses were received from eight of these states. Although there were practical constraints on collecting this data due to limited time, certain details were omitted from this early draft of the paper which would have been desirable from a methodological perspective. For example, it is not entirely clear exactly what the state department representatives were asked in terms of survey items. In addition, information about sampling would have been useful as well. For example, how were these seventeen states selected? In examining the states that responded, some of the states with significant numbers of language minority students were absent, including Florida, California, and Arizona. Since states vary significantly regarding proportions of language minority students, they are not all weighted equally in terms of importance, and it would be interesting to have additional data on what other states are doing. These limitations in terms of sampling need to be taken into account in interpreting the generalizability of the survey results.

Notwithstanding these potential limitations regarding generalizability, there appear to be two major findings which emerged from the survey. First, there is wide variation in terms of current state practices. This is not altogether surprising, however, it does suggest that aggregating data and arriving at a summary statement regarding national practices is not a simple or direct matter.

The second finding is that, if the measures the states are using are examined, in addition to the Home Language Survey which is used by almost all, there is a strong reliance on standardized tests and oral language proficiency tests. In attempting to evaluate this pattern, it is useful to ask what is currently known about language in terms of research and theory and then compare that with current practices.

Although the body of research on language and bilingualism is immense and complex, there are some generalizations which would likely result in wide agreement. For example, work in linguistics, anthropology, cross-cultural psychology, cognitive psychology, and other fields suggest that language use (and by extension, "proficiency") is context-sensitive and context-specific. Proficiency is no longer viewed as a fixed, invariant, "within-the-head" phenomenon. Secondly, language is inherently social. It is acquired and used in social settings for social purposes. Thirdly, language is acquired, not learned. That is, it is rare to see a parent saying to a child, "Today we are going to learn plurals" in the normal course of the day's activities. It is acquired in natural settings in the course of people's needs to accomplish specific social activities such as eating, dressing, and so forth.

Another thing which is known about language is that it is used in order to accomplish meaningful activities. That is, it is purposeful, a tool in order to accomplish everyday tasks. It is also a tool in the sense of being a sign system, which is used to mediate cognition. In this sense, there is an intriguing link between language and thought as Piaget, Vygotsky, and many others have noted. A final point about language is that it can be seen as an integrated part of a larger system of literacy. Therefore, if language is broadened to include written language and so forth, then perhaps the focus on oral language is overly narrow.

If the above generalizations about language are taken as a simplified summary of current views, and compared to the reported practices of state departments of education, there is not a great deal of correspondence or match. Specifically, the heavy reliance on standardized tests, achievement tests, and oral language proficiency measures as reported in the survey suggest that an outdated view of language is being used to drive practice.

Comments on an Alternative Model: The ELAN Profile Chart

Taking the same general points about language as a starting point, the author's proposed model can be compared to the generalizations described above as well. In the paper, Dr. Canales discussed the theoretical base of the model as being sociolinguistic. As described in the paper, it is fairly consistent with the generalizations of language outlined earlier, certainly much closer than reported school practices.

The proposed model suggests that data be collected in three areas: oral language proficiency, social data, and academic data. The use of multiple evaluative criteria is suggested, and it is proposed that the scores can then be converted to Likert-scale ratings. From these converted ratings, a profile can be constructed and a classification derived, resulting in a designation of either advanced, intermediate, or beginning level.

Although the proposed model is certainly more comprehensive than what is currently being carried in many school districts, there are some components which might merit consideration for inclusion. One, for example, would be data on the affective state of the child with respect to first and second languages and their usage. As my colleague at USC, Steve Krashen suggests, the affective state of the child is important in terms of how rapid and effective the second language acquisition process is, and is an additional but important piece of data.

Another important piece of data of great interest would be the socio-political context in which the first and second languages are being or have been acquired. The relative status of L1 and L2 has an important impact on the child's acquisition of language, yet it is normally ignored in the assessment process because the focus is exclusively on the child.

A major component of the ELAN Profile Chart is the Likert-scale score conversions, which in essence is a data-reduction technique. That is, data from various types of proposed measures are converted to a five-point scale, making the data more comparable. However, when data is reduced by this or any other technique, precision is lost. As an example, a percentile score of 83.5 on a standardized measure, when converted to its transformed equivalent on a five point scale in order to make it more comparable to other data, loses some precision. This may be useful in aggregating and summarizing data across districts and/or states, however data is converted to an ordinal scale of measurement. That is, it is possible to say that a four is less than a five, but not how much more, and the distance between a three and a four, for example, may not be equivalent to the distance between a four and a five.

Another consideration in the proposed model is that equal weights are given to each of the proposed indicators, if my understanding is correct. Assuming that it is, this would suggest that the data from the Home Language Survey would be equivalent in importance to protracted observational data in a large number of contexts. Is it logical to equate the meaningfulness or usefulness of these distinctly different sources of data? I would suggest that this point is certainly open to question.

One of the curious aspects of the proposed ELAN Profile Chart is that many of the alternative measures proposed were already listed as options by many state departments of education. An important question, it seems, is why are states not using these measures already? These alternatives to standardized tests already exist and are available, suggesting that perhaps the development of completely new measures may not be what is needed in the assessment of LEP students. The alternatives which do exist are not extensively used, and I will return to this point shortly.

One point of contention with the proposed ELAN model would be the almost exclusive focus on oral language, more specifically English oral language. It seems this is overly restrictive in light of how language and literacy are currently viewed. From my perspective, it would be desirable to consider relative linguistic proficiency, not only in English but in the child's native language as well. Secondly, I would suggest broadening the scope to a wider focus on literacy as opposed to oral language exclusively. This might mean more attention to written language and other forms of literacy which are traditionally separated for assessment and instructional purposes. However, given the strong relationships among these, and the current view of language and literacy as part of a complex whole, separating out oral language from other parts of the child's development may not be the most advisable course.

A final point with respect to the ELAN model has to do with the distinction between classification and diagnosis. The former is the term for sorting and comparing students. That is, who is lower? Who is higher? Who goes into this group? Who goes into that group? The latter term, in contrast, refers to data used to derive intervention or treatment. The conceptual distinction between these two terms is often confused in discussions or assessment procedures. My understanding of the ELAN model suggests that it is concerned with the issue of classification. Certainly Likert-scale conversions will allow one to say who is higher and who is lower on one or more measures. However, data of this type are not terribly useful for day-to-day instructional decisions. Data of the type provided by converted standardized scores are severely limited. If the concern is "What does this child know and what is the next thing this child needs to work on?" For the practitioner needing to know, "What do I do with this particular child today?" Global comparative data does not provide a very specific answer. Simply put, I would like to argue for increased attention to instructional relevance and data more accessible to instructional personnel.

Considerations for Future Revisions

Obstacles to change. In this final section, I would like to provide some suggestions for consideration in future revisions to the paper presented. One critical question has to do with the obstacles to change in educational institutions. A great deal of attention is currently being given in assessment circles to alternatives to traditional standardized assessment, which many have described as problematic. Why is it, however, that even when alternatives are available they are not heavily used? I would like to propose two hypotheses which might merit consideration as alternative assessment models are developed and considered.

One hypothesis is that teachers, bilingual specialists, and other practitioners in school settings have a particular schema or mental model of assessment. That is, this mental model provides a unified, logical framework of thinking about what assessment is, why it is used, how it fits together with instruction, and so forth. One possibility is that the mental model of assessment embedded in schools is very different from that embedded in the work of those researchers and theoreticians concerned with developing alternative assessment models. However, these underlying assumptions and belief systems are rarely taken into account. Innovative practices which do not neatly fit into one's existing mental model are ignored or discarded. Simply put, it is not enough to develop and disseminate alternative assessment models or procedures without taking into account the existing belief structures of the "end users." When viewed in this perspective, the failure of school practitioners to incorporate new assessment developments is logical and understandable. Unfortunately, rather than examining test-users, research (mostly guided by a psychometric framework), has tended to concentrate on the technical characteristics or procedural aspects of the tests themselves with little attention to those who would use them. It is important to recognize that many of the new innovations in assessment methodology and theory are rooted in a different paradigmatic framework from that familiar to many practitioners.

A second hypothesis is based on Mehan's work on educational decision making in special education. In his ethnographic examination of the referral, assessment, and placement process, he found that decisions were rarely made on the basis of rational consideration of test data and other child-related characteristics, as is assumed to take place in current law. Rather, he found the process to be characterized by "social negotiation," trade-offs, and bargaining; A child's educational fate often depended upon these interpersonal negotiations among educational personnel. In trying to make sense of these findings, Mehan assumed that all the actors were not malicious or incompetent. Rather, he concluded that their behavior was rational given the "institutional constraints" under which they were forced to operate: limited time, budgetary shortages, conflicting laws, and so forth. The conclusion was that these very powerful everyday constraints had an overwhelming impact on day-to-day behavior, and what appeared irrational on the surface actually made sense. By extension, it can be assumed that there are such constraints in institutional settings such as state departments of education, school districts, and individual classrooms which mitigate against change. These have yet to be studied, although it is possible that they exert significant pressure on the implementation of new assessment procedures.

The Larger Context of Assessment

One point that I would like to see addressed in this paper is increased consideration of recent developments regarding assessment at the national level. As an example, there is much talk about more authentic assessment to reflect closer alignment to authentic curriculum (c.f., the California Language Arts Framework) and to recent theories of cognition and learning. Portfolios and other innovations are being widely discussed, even as pressure is mounting for national indicators of performance. It is likely that the next few years may usher in significant change in how assessment is conceptualized and used because of events taking place at the national level. The work discussed in this present paper under consideration should not be treated in isolation from these developments, but rather should be considered within that larger context.

The issue of entry and exit. One factor which might merit further attention in future work on this topic is the whole issue of entry to and exit from bilingual programs. At present, it appears that schools operate from a rather inflexible, all-or-none system that is heavily reliant on standardized assessments. It would be useful to consider more flexibility within this system, especially since learning is not conceptualized in such an all-or-none fashion. How could alternative assessment for LEP students be restructured to assist in this process?

The issue of eligibility. Because of my background in special education, I have a special sensitivity to the whole issue of eligibility. This has been a central concern of the field, and I would like to hope that in the treatment of language minority students we learn from the mistakes which have been made. Historically, much attention has been placed on the question, Who has learning problems, and who does not? Who should receive services and who should be excluded? Tremendous amounts of scarce resources are spent on generating psychological reports and making complicated eligibility determinations. Entry into the system in most cases is dependent upon meeting a certain profile or criteria. In spite of the fact that the assessment methodology and procedures are often technically inadequate, the field has focused on making finer and finer distinctions between groups of students at a tremendous cost. However, much of the assessment data collected during this sorting process does not readily translate into educational prescriptions. Moreover, many have argued that there are not really separate treatments for all the various diagnostic categories once they are filled.

The field of special education is currently in the midst of widespread controversy precisely because of these factors. It would be my hope that, in the field of bilingual education, we could avoid and even learn from some of these same mistakes. In order to meet these challenges, truly innovative developments are required at the level of assessment. However, it is not sufficient to consider the procedural aspects of new assessment methodologies apart from the new paradigms in which they are embedded or apart from the social contexts in which they will be used, that is, individual classrooms.

Test Score Pollution: Implications for Limited English Proficient Students

Thomas Haladyna
Arizona State University, Tempe

Introduction

Standardized tests have a multitude of interpretations and uses. Test score pollution is a condition that affects the validity of these interpretations and uses. This paper presents the problem of test score pollution in the context of achievement testing, speculates about its origins, provides evidence of its complexity and severity, and addresses the implications of test score pollution for limited English proficient students.

Test Score Pollution: Implications for Limited English Proficient Students

Current reform in the organization of schooling has been accompanied by significant reform in testing (Toch, 1991). Standardized achievement tests have been under siege for many years (Hoffman, 1964; Fair Test Examiner, 1987), and "authentic assessment" has recently been proposed as an alternative or replacement for the standardized achievement test. Baker (1991) summarized the prevailing attitude behind this test reform when she stated that the authentic assessment is more holistic and realistic of what real teaching represents, while the standardized testing is more molecular and facts-based.

Part of the testing reform movement can be attributed to persistent criticism that standardized achievement tests fail to measure the important outcomes of schooling or that it only partially measures these outcomes (Berk, 1988; Brandt, 1989; Frederiksen, 1984; Haertel 1986; Haertel and Calfee, 1983; Linn, 1987; Madaus, 1988; Messick, 1987; Shepard, 1989).

The topic of this paper is the second of a two-faceted problem involving achievement testing in the United States. The first facet is the lack of correspondence between test content and intended student outcomes in school districts, and the second facet is "**test score pollution.**" This term describes instances where test scores for a unit of analysis (such as a class or school) are systematically inflated or deflated without corresponding changes in the content domain that a test is supposed to represent (Haladyna, Nolen, and Haas, 1991). Whether we use a standardized test or an authentic, assessment is probably irrelevant. Because standardized achievement tests have been used for many years, test score pollution is associated with this type of test, but authentic assessments may be even more susceptible to test score pollution (Canner, 1991).

First, we examine the concept of validity. Second, we look carefully at the meaning of school achievement. Third, we define test score pollution and then evaluate the research bearing on this problem, and finally we speculate about the effects of test score pollution on limited English proficient (LEP) students.

Construct Validity

Traditionally the topic of validity has been treated in three categories (construct, criterion-related, and content), but recently Messick (1989) has presented a unified approach to validity under the rubric "construct validity." In this conceptualization, validity refers to interpretations as well as uses of test results.

For instance, Haladyna, et al. (1991) presented 29 different uses of standardized achievement test scores. Table 1 summarizes these interpretations and uses. Dorr-Bremme and Herman (1986) offer findings from their national survey illustrating the variety of uses of test results.

Table 1
Consumers and Uses of Standardized Achievement Test Information

Consumer: National Level	***Units of Analysis***
Allocation of Resources to Programs and Priorities	Nations, States
Federal Program Evaluation (e.g., Chapter 1)	States, Programs
Consumer: State Legislature/State Department of Education	
Evaluate State's Status and Progress Relevant to Standards	State
State Program Evaluation	State, Program
Allocation of Resources	Districts, Schools

Consumer: Public (Lay persons, Press, School Board Members, Parents)

Evaluate State's Status and Progress Relevant to Standards	Districts
Diagnose Achievement Deficits	Individual, Schools
Develop Expectations for Future Success in School	Individuals

Consumer: School Districts--Central Administrators

Evaluate Districts	Districts
Evaluate Schools	Schools
Evaluate Teachers	Classrooms
Evaluate Curriculum	District
Evaluate Instructional Programs	Programs
Determine Areas for Revision of Curriculum and Instruction	District

Consumer: School Districts--Building Administrators

Evaluate School	School
Evaluate Teacher	Classrooms
Grouping Students for Instruction	Individuals
Placement into Special Programs	Programs

Consumer: School Districts--Teachers

Grouping Students for Instruction	Individuals
Evaluating and Planning the Curriculum	Classroom
Evaluating and Planning Instruction	Classroom
Evaluating Teaching	Classroom
Diagnosing Achievement Deficits	Classroom, Individuals
Promotion and Graduation	Individuals
Placement into Special Programs (e.g., Gifted, Handicapped)	Individuals

Consumer: Educational Laboratories, Centers, Universities

Policy Analysis	All units
Evaluation Studies	All units
Other Applied Research	All units
Basic Research	All units

While many observers do not support these interpretations and uses, little doubt should exist that researchers, evaluators, policy analysts, and lay persons (including legislators and the press) are interested in interpreting and using test results in these ways.

The Standards for Educational and Psychological Testing (American Psychological Association, 1985) are very explicit about the need to validate any interpretation or use. Standard 1.1 on page 13 states:

> *"Evidence of validity should be presented for the major types of inferences for which the use of a test is recommended. A rationale should be provided to support the particular mix of evidence presented for intended uses."*

In a national survey by Hall and Kleine (1990), 90 percent of the respondents reported that tests are used to evaluate teacher effectiveness. Berk (1989) and Haertel (1986) have offered strong criticism against such use. Another example is the use of state-by-state comparisons to draw inferences about a state's success at educating its students, a practice that has received much criticism (Guskey, & Kifer, 1990; Koretz, 1991).

A storm of protest about the misinterpretation and misuse of test scores has existed for years within the community of testing specialists education (e.g., Brandt, 1987; Frederiksen, 1984; Haertel, 1986; Haertel and Calfee, 1983; Linn, 1987; Madaus, 1988; Messick, 1987; Shepard, 1989). As test users, we must be vigilant about misinterpretation and misuse of test results for purposes of evaluation and policy making affecting our jurisdictions.

Construct validation calls for the collecting of evidence to support any of the 29 different uses or interpretations of test results that we desire. Messick (1989) provides a very comprehensive discussion of construct validation and the logical and empirical types of evidence necessary to validate test interpretations and uses. Without such evidence, we should question the ethics of those within the profession of education making unsupported claims based upon test results. Seldom do we see evidence presented to support any of the interpretations and uses found in Table 1. Consequently, we should resist attempts to interpret or use test results in ways unintended and unsupported by validating evidence.

School Achievement

School achievement is the main construct of education. Hypothetically, we can define school achievement in terms of many subject matter areas, using instructional objectives, and organize these objectives by content and by a level of cognitive behavior, such as found in the Bloom taxonomy. An explicit, national curriculum does not exist, but the belief that the standardized achievement test reflects this general national curriculum has been expressed at various times by various writers (e.g., Freeman, Belli, Porter, Floden, Schmidt, & Schwille, 1983; Leinhardt & Seewald, 1981; Phillips & Mehrens, 1987). In general mixed evidence exists on this issue of whether the test represents a national curriculum, but staunch advocates of systematic instruction argue that no standardized achievement test is likely to be interchangeable and represents specific classrooms, curricula, and instruction (Cohen, 1987; Nitko, 1989).

The Arizona Department of Education learned recently that only about 27 percent of its essential skills could be found on a standardized achievement test (Noggle, 1988). The Department of Education changed its testing program to provide a closer alignment to its state-mandated essential skills curriculum. Other states, like Missouri, have already accomplished this. School achievement is going to have to be redefined by a jurisdiction, and carefully measured, if reform in testing is to be effective.

Several researchers have questioned the kinds of inferences we can draw from standardized achievement test data (Nolet and Tindal, 1990; Wardrop, Anderson, Hively, Hastings, Anderson, and Muller, 1982). They claim that only general interpretations can be made about standardized achievement test results. Test companies have never claimed that their tests measure school curricula, instructional practices in school districts, schools, or classrooms (Mehrens & Kaminski, 1989). Koretz (1989, p. 33) stated it succinctly:

> *"Put simply, an achievement test is typically a brief and incomplete proxy for a more comprehensive, but less practical, assessment of some domain of achievement."*

Teachers generally believe that standardized test results do not reflect their teaching and they tend to rely on their own observations (Dorr-Bremme & Herman, 1986; Haas, et al., 1989).

Causal Attribution

Part of the problem of achievement is the strong desire to know what or who has caused students to achieve or not achieve. Accountability requires that we make causal statements about achievement. School achievement is the result of many influences existing over a child's lifetime and even prior to a child's birth. Some of these factors, such as family and home influences, parental education, socioeconomic status, family mobility, and neighborhood exist outside the influence of schooling. Other factors, such as learning environment, motivation and attitude, and quality and quantity of instruction, are under the influence of school personnel. While we have trouble measuring school achievement, we have even more trouble with causal attribution. We have not yet completely understood the influence and interactions of these variables on school learning, although models like Walberg's productivity model (Walberg, 1980) provide a workable framework for our understanding of causes of learning. Lay persons tend to oversimplify education by using test results as the operational definition of achievement and the teacher as the singular cause of school learning.

Higher Level Thinking

A common distinction among all educators is that student learning comes in various forms of mental complexity, ranging from recall to various types of higher level thinking, often expressed in the Bloom taxonomy. Many critics and researchers alike have concluded that curricula, teaching, and testing have focused on lower level thinking, such as recall, at the expense of hard-to-measure higher level thinking outcomes. Nickerson (1989) leaves little doubt that American education will focus on making its students thinkers, and therefore higher level thinking will become a strong feature of new standardized achievement tests.

A dilemma presents itself (Haas, Haladyna, and Nolen, 1990; Nolen, Haladyna, and Haas, in press; Smith, 1991): Teachers are forced to give standardized tests, which they believe measure lower level thinking. Some teachers promote higher level thinking in their classrooms at the expense of preparing students for the standardized tests, while other teachers faithfully drill students on the kinds of outcomes known to be tested. Who is the more effective teacher? This dilemma is part of the problem of test score pollution.

The problem of testing higher level thinking is further complicated by recent reports that teachers are either reluctant or unable to develop classroom tests to measure higher level thinking (e.g., Stiggins, Griswold, & Wikelund, 1989), while standardized tests are equally at fault for failing to measure higher level thinking. Nonetheless, the new thrust in performance testing (euphemistically referred to as "authentic assessment") promises to give greater emphasis to the measurement of higher level thinking through the development of multi-step exercises.

Multiple-Choice versus Performance

A current opinion held in education is that performance tests measure higher level thinking outcomes while multiple-choice tests measure recall, and other trivial forms of behavior (Baker, 1991).

Recent and past reviews of research on the equivalence of open-ended versus selected-response formats reveals their equivalence (Bennett, Rock, and Wang (1990). Further these researchers submit that the stereotype that multiple-choice tests measure trivial content and factual recall while open-ended tests measure higher level thinking is FALSE.

Measurement specialists have consistently maintained that multiple-choice items can be used to measure higher level thinking outcomes, admitting that it is difficult to do via any format. For instance, the context-dependent item set that contains a stimulus and a set of test questions can be used to measure various types of higher level thinking outcomes via a multiple-choice format (Haladyna, 1991, in press a, in press b).

Conclusion

School achievement is a complex constellation of knowledge and skill that is difficult if not impossible to measure with a single test. Therefore, no current test seems to be adequate toward the end of measuring the complete domain represented by a school district's curriculum. Further, we lack many technologies in item writing and scoring to measure adequately many aspects of human behavior.

The variety of purposes listed in Table 1 are not served by using a standardized achievement test. That is why many observers call for significant reform in testing where multiple indicators are used and where achievement is better defined in terms of its many aspects.

Test Score Pollution

Test score pollution is any influence that affects the accuracy of achievement test scores. Messick (1984) called these influences "contaminants" but did not specify exactly what these contaminants are. Haladyna, Nolen, and Haas (1990) identified three sources of contamination and reviewed the research bearing the seriousness of each. These are: (1) test preparation, (2) situational factors, and (3) external conditions. Table 2 provides a list of 21 specific sources of test score pollution organized by these three categories, adopted from Haladyna *et al.* (1991).

Table 2
21 Documented Sources of Test Score Pollution

Test Preparation Activities

Testwiseness Training
Increasing Motivation
Curriculum Matching
Changes in the Instructional Program
Specific Inappropriate Instruction (*Scoring High*)
Presenting Items Similar to Those Found on the Test
Presenting Items Identical to Those Found on the Test
Excusing Low-achieving Students From Taking the Test
Cheating

Situational Factors

Test Anxiety
Stress
Fatigue
Speededness of the Test
Motivation
Recopying and Checking Answer Sheets
Test Administration Practices

Context

Language Deficits
Socioeconomic Context
Family Mobility
Family and Home Influences
Prenatal/Early Infant Influences

Origins of Test Score Pollution

Undoubtedly, the range of uses of standardized test scores has changed drastically from the 1950s to the 1990s (Haertel and Calfee, 1983). The current overuse and misuse of test results, coupled with the "high stakes[1]" nature of many uses has badgered superintendent, principals, and teachers to prepare students to perform on these tests. According to Haas et al. (1990), although the preparation forces teachers to depart from regular instructional practices and teachers almost uniformly dislike the test and disagree with the public's misuse of test results, the pressure to produce high test scores is unbearable. One teacher commented:

> *...I feel that if I am pressured any more to do well on the TEST, I will do everything I can to make sure my kids do well...even cheat. I have a family to support and I would be stupid not to do this. My job is more important than my values.* (Haas, et al., 1990, p. 128).

Test Preparation

A variety of school activities falls into the category of test preparation. Haladyna et al., (1990), Mehrens and Kaminski, 1989) and Smith (1991) present a continuum of test preparation activities. The following is Smith's conceptualization.

The first is **no special preparation**. Nolen et al., (in press) reported that 12 percent of teachers surveyed did no special preparation. The fact that 88 percent did introduces a form of pollution.

The second is to **teach test-taking skills**. Nolen et al., (in press) reported that over 60 percent of teachers surveyed did this. Test taking skills (or "testwiseness" as it is sometimes referred to) is well defined in the extant literature, and Bangert-Drowns, Kulik, and Kulik (1983) and (Sarnacki (1979) reported that indeed testwiseness training does work. Comparisons between those teaching test-taking skills and those not teaching test-taking skills introduce test score pollution.

A third method is **exhortation**. This includes advice on eating and sleeping before the test, pep rallies, the principal's announcements and words of encouragement, and other measures designed to "motivate" students to do their best on the "test."

A fourth method is the **design of instruction to match the test content**. Some materials, such as *Scoring High in Math* (Foreman & Kaplan, 1986), appear designed to identify the exact content of a standardized test and to provide specific instruction on this material (Mehrens & Kaminski, 1989). Toch (1991) presents a more comprehensive description of the extent of the industry for producing materials to prepare for standardized achievement tests. Haas *et al.* (1990), Nolen et al., (in press) and Smith, Edelsky, Draper, Rottenberg, and Cherland (1989) report extensive use of these materials in elementary school classrooms as well as disenchantment with this practice. A national survey conducted by Hall and Kleine (1990) revealed that 69 percent of the sample reported changes in the curriculum to match the standardized achievement test, 39 percent reported changes in the curriculum to match particular questions on these tests, and 82 percent reported teaching material because it is on the test. Several critics of these practices have stated that the curriculum, in effect, is narrowed, that time for instruction on non-test related and other important content is lost, that instruction is very test like, and that both teachers and students suffer in many ways (Smith & Rottenberg, in press). Popham (1990), among others, criticized the ethics of this narrowing of curriculum and instruction.

A fifth method is "**stress inoculation**." Teachers report helping students boost test scores for the purpose of increasing the students' collective self-respect. Since the improvement or maintenance of self-respect is so important, the achievement of high test scores is viewed as a vehicle for this worthy goal.

A sixth method is **practicing on items of the test itself or a parallel form**. Both Nolen, et al., (in press) and Mehrens and Kaminski (1989) stated that about 10 percent of teachers reported doing this. While these researchers believe that this is blatantly dishonest, some teachers believe that since the tests are so inherently misused and misinterpreted, this practice is done to "play the game" with administration and the school board.

A seventh method, **cheating,** refers to giving answers to students, providing hints to students, and changing answer sheets after the test.

Table 3 provides a list of test preparation activities from Haladyna, et al., (1991), and their judgments regarding how ethical these test preparation practices are. Mehrens and Kaminski (1989) offer a similar set of judgments, and Cannell (1988) also provides his appraisal of the ethics of various test preparation practices. Haladyna et al., (1990) also make the point that despite whether a test preparation activity is ethical or not, all test preparation activities are polluting if one class, school, or school district does it while others do not.

Table 3
A Continuum of Test Preparation Activities

Test Preparation Activity:	Ethical Degree
Training in testwiseness skills	Ethical
Checking answer sheets to make sure that each has been properly completed.	Ethical[1]
Increasing student motivation to perform on the test through appeals to parents, students, and teachers.	Ethical
Developing a curriculum based on the content of the test.	Unethical
Preparing objectives based on items on the test and teaching accordingly.	Unethical
Presenting items similar to those on the test.	Unethical
Using *Scoring High* or other score-boosting activities.	Unethical
Dismissing low-achieving students on testing day to artificially boost test scores.	Highly Unethical
Presenting items verbatim from the test to be given.	Highly Unethical

[1] Ethical to the extent that the test publisher recommends it or to the extent that all schools, classes, and students being compared have the same service.

Another aspect of undesirable test preparation is that by raising test scores, there is no correlated gain in the general domain of achievement that each test is supposed to represent. Recently, Koretz (1991) presented some evidence to support this suspicion, and more research results are expected to further support the polluting influence of many forms of test preparation. Linn Graue, and Sanders (1990) concur with Cannell's findings (Cannell, 1988), that achievement scores are higher than ever, but they assert that the problem may indicate (1) teaching too specifically to the test while at the same time the norms are not keeping up with this specific form and (2) questionable forms of test preparation.

Situational Factors

Haladyna, et al., (1990) in their review of research on test score pollution have documented many factors that are specific to the administration of the test and are also very polluting. Some of these may have saliency for LEP students and these will be addressed more fully in another section of this paper.

Test anxiety. Kennedy Hill and his colleagues (Hill, 1979; Hill & Wigfield, 1984; Hill & Sarason, 1966) have extensively studied test anxiety and estimate that over 25 percent of the school age population have some debilitating form of this disorder. Test anxiety is treatable, but it is also exacerbated by stress-producing conditions in the classroom and school. If an explicit or implied threat exists, test anxiety can be increased (Zatz and Chassin, 1985). Mine, and others (1987) noted that some Japanese families actually promote high test anxiety through parental restriction, blame, inconsistency, overprotection, and rejection. They also state that praise has the *same* effect on test anxiety instead of the opposite effect.

Stress. Children experience many stress-provoking situations in life, many of which are related to school or affect school life (Karr and Johnson, 1987). Oddly, little is known about stress in the classroom. Recent reports give some credence to the role of stress in standardized testing situations (e.g., Nolen, et al., in press; Paris, Lawton, Turner, & Roth, 1991).

In the Paris et al., study, they specifically asked children questions about the effects of the testing experience. Three aspects of why stress may be increased under the condition of the standardized testing experience are that (1) students become increasingly skeptical about the value of test results as they become older, (2) the purposes or uses of the test are not clearly revealed, (3) there is a social impact on students based on their test score status.

Fatigue. Reports of fatigue during the testing process, particularly with younger children, have been reported (Dorr-Bremme & Herman, 1986; Haas et al., 1990; Nolen, et al., in press; Smith et al., 1989). In sun belt states, such as Arizona, temperatures during May testing may reach into the 90s or low 100s, a condition that increases this potential source of pollution. Interestingly, there is no research that specifically addresses the problem of test fatigue.

Timed testing. One condition of all standardized tests of this type is the time limit, which must be strictly followed to provide standardized test results. Reports of plodders and sprinters in timed tests reveal a possible source of test score pollution (Wright and Stone, 1979). This factor is particularly significant to LEP learners and, it will be treated more extensively in another section of this paper. In addition, timed testing seems particularly harmful to test anxious children (Plass and Hill, 1986). Wodtke, Harper, Schommer, and Brunellia (1990) report liberal violations of time limits in tests administered by teachers. Hall and Kleine (1990) reported that 9 percent of the teachers surveyed in their national study felt pressured to extend time limits and commit other nonstandard testing practices. If the stakes for test results are indeed very high, this should come as no surprise.

"Blowing off the test." Motivation to perform on the test is very important to test performance. Some school districts expend considerable effort in motivating its students, while other districts do not. Haladyna, et al., (1990) identify a host of factors known to increase or decrease performance, all of which are in some way related to motivation. Widespread reports exist that younger students are likely to be more attentive to the test but that older students, seeing the lack of consequence for their test performance, will often resort to random marking (Paris, Turner, & Lawton, 1990). Dorr-Bremme (1986) also reported anecdotal evidence from interviews suggesting that many students do not give much effort to performing well on these tests.

Teacher attitudes may have something to do with test performance. When teachers are highly motivated to get high test scores, student performance may be maximal. With poorly motivated teachers, students merely go through the motions, knowing that the results mean nothing to the teacher. While this hypothesis about teacher attitude is very speculative, anecdotal reports in Haas, et al., (1990) reveal widespread discontent with the standardized test and with the motivation of students to perform on these tests. Smith (1991) also discusses the discouraging climate that standardized testing creates for teachers and the dilution of their professionalism.

Recopying, checking, and repairing mismarked answer sheets. Some school districts have policies that allow the checking of answer sheets for stray marks and light marks, or mismarked answers. Parents, other volunteers, or paid classroom aides are asked to check answer sheets in some schools. The fact that some schools or districts have policies and procedures for this practice while others do not creates another possible source of pollution.

Summary. This section has provided a brief overview of possible test score polluting practices that reside in the test administration or events preceding test administration that do not include test preparation. While many of these practices exist in schools, we know very little about the importance of each as a test score pollutant. Still, indications from this limited research suggest that our concern is warranted and further study is needed.

External Factors

Anyone close to the educational process knows the many factors that underlie poor test performance: inadequate prenatal care, low mental ability, poor early childhood nutrition, lack of social capital in the family and home, disintegrating family social structure, poor motivation, LEP, low socioeconomic status, high family mobility, and lack of education of parents. While this list is brief and hardly all inclusive, it represents factors *outside* the influence of schools and school personnel that are believed to affect school performance. In various evaluation and policy studies at national, state, and school district levels, seldom is reference given to the influence of these variables on test scores. In actuality, schools and school personnel are often given the "blame" or "praise" for test scores that were obviously influenced by these external factors. Therefore, these factors, when unnoticed or not considered, are a source of test score pollution because they affect the accuracy of test score interpretations and uses.

Acting on a state law, Arizona's Department of Education has to report all standardized test scores in the context of two external factors, language proficiency and socioeconomic status (as determined by frequency of use of the school lunch program). Model reporting systems such as this one attempt to reduce the severity of pollution from these external factors.

Implications for Limited English Proficient Children

This section of the paper addresses implications for LEP educators arising from the problem of test score pollution. This section also suggests some fruitful areas for research on the role or influence of test score pollution on LEP students. Finally, recommendations are offered to protect LEP children from negative consequences due to using polluted test scores.

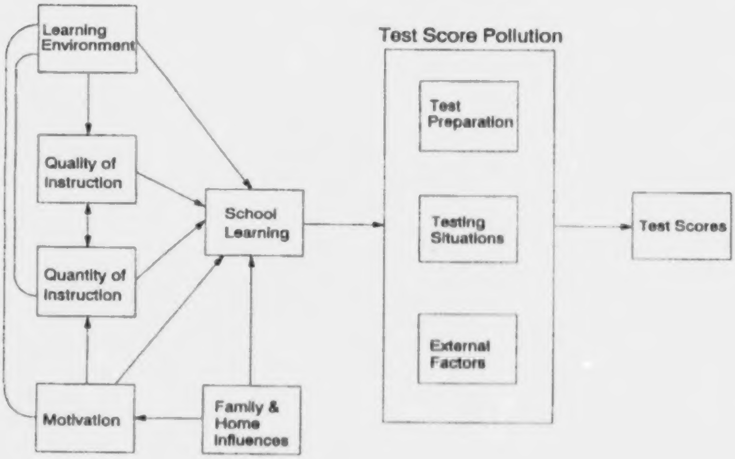
This section of the paper is loosely based on a working model of school learning that includes test score pollution. The following review of research is not very comprehensive but helps build a working hypothesis about why we should be very cautious about test scores obtained from LEP children.

A Causal Model of School Learning Modified to Accommodate Test Score Pollution

To begin this section, a causal model of test performance is offered that is loosely based on the Walberg productivity model

(Walberg, 1980). Figure 1 provides an illustration of the model. The elements are familiar to most educators, and various studies and meta-analyses speak of the potential influence of such constructs as family and home as causal determinants of children's motivation and their learning--as inferred from a non polluted standardized achievement test. Quality and quantity of schooling is also positively and causally related to learning. Learning environment contributes to a high quality of instruction and increases learning time, quantity of instruction, which, in turn, leads to better learning. Learning is demonstrated in many ways in schools, grades being one indicator. The standardized achievement test, at best, provides a gross, general measure of school learning (Nolet and Tindal, 1990; Wardrop, Anderson, Hively, Hastings, Anderson, and Muller, 1982), but as Figure 1 shows, all test performance is mediated by the three possible forms of test score pollution. Therefore, no test score interpretation or use, for any unit of analysis (class, school, district, state, or nation) is valid until we can eliminate the influences of test score pollution.

Figure 1
The Role of Test Score Pollution in Interpreting School Achievement



Facts About LEP Children

As a prelude to the following discussion, several facts about LEP children should be stated. For instance, in a recent publication from the National Center for Education Statistics (Rock, Pollack, & Hafner, 1991), the performances of LEP children as well as other demographics are well documented.

First, and most obvious, LEP children have the handicap of reading, writing, speaking, and listening in a foreign language. Levels of facility in English vary and handicap these children's test performance. Another source of evidence comes from Arizona state testing (Bishop, 1988), which contains information about the test performances of LEP and English proficient children in Arizona. The typical range of LEP children's performance on the state's mandated standardized achievement test ranges between the 14th and 43rd percentiles, while the English primary language students' performance level is near the 62nd percentile. Rock, et al., (1991) report from their national sample of LEP and non LEP students in reading, mathematics, science, and history/citizenship/government that language facility is indeed an important factor in test performance. Effect sizes ranged from .58 for reading to 1.07 for the social studies factor. These are substantial differences.

Second, most LEP children are below average in terms of socio economic status.

Third, the majority of LEP children are from ethnic groups, and each has its distinct culture (Rock et. al., 1991). More than one half of the LEP children in their national sample are Spanish-speaking, and they are more handicapped than those LEP children who speak other languages.

Fourth, LEP education programs offer a "non mainstream" experience designed to help LEP students become mainstream students, but the process of being in LEP programs socially distinguishes these students from mainstream students in social and intellectual ways.

If these assumptions are tenable, the following review of research and discussion bears on test score pollution for LEP children.

Standards

The Standards for Educational and Psychological Testing (American Psychological Association, 1985) are explicitly concerned about LEP students, and it seems worthwhile to review several standards in relation to this problem of test score pollution. Standard 13.1 (page 74) states:

> *"For non-native English speakers or for speakers of some dialects of English, testing should be designed to minimize threats to test reliability and validity that may arise from language differences."*

Studies cited in the next section of this paper give some evidence for potential bias against LEP students. Standard 13.3 (p. 75) states:

> *When a test is recommended for use with linguistically diverse test takers, test developers and publishers should provide the information necessary for appropriate test use and interpretation.*

If the test manual lacks this information, we should submit that the test is probably NOT suitable for LEP persons, since the potential for polluted test scores is too great to risk using the score for any important educational decision. Standard 13.5 (p. 75) states:

> *"In employment, licensing, and certification testing, the English language proficiency level of the test should not exceed that appropriate to the relevant occupation profession."*

This is a serious threat to the validity of professional licensing examinations and tests used to make personnel decisions. Since LEP persons typically have a significant handicap in reading, the existence of unnecessarily difficult reading levels in "high-stakes" tests creates a significant yet subtle form of bias. It would be easy to challenge an examination that has high reading demand on examinees as an example of adverse impact on LEP students.

Test Interpretations and Use

As Table 1 attests, we have witnessed a steady increase in the number and variety of interpretations and uses of achievement test scores. The issue is validity. Some of these interpretations and uses have serious consequences on the extent of education and futures of all children. For instance, test scores are used for placement into special programs (for handicapped or gifted) and for placement in achievement tracks (for example, in courses ranging from beginning to advanced mathematics). Such tests are also used for minimum competency decisions, for example, for high school graduation or promotion.

The first point about test interpretation and use is that it behooves test users to ensure that these scores are unpolluted before using test results. A second point is that the placement of children in programs strictly based on test scores should be questioned. If LEP children's test performances are lower due to test score pollution, then the system that misuses these scores for these various assignments is at fault.

Test Preparation

All children should be experienced test takers. They should have comprehensive test-taking courses and be equally skilled in test-taking. Popham (1990) also submits that practice testing on content related to the test is reasonable if the test formats are varied to encompass a wide range of possible test formats, since focused practice on the actual format of the test may lead to spuriously high results.

Since LEP students typically lack testing experience of this type, they also may lack test-taking skills. Without the experience of test-taking coupled with test-taking skills, they suffer a significant handicap. This inexperience may contribute to other test pollution problems, such as test anxiety. All other forms of test preparation should be viewed as contradictory to effective teaching and fair uses of standardized test results. Any attempt to promote high test performance through other means should be viewed the way the public views the use of steroids for body building, a dangerous and unhealthy shortcut. Moreover, the spurious increase in test performance due to these test preparation activities does not represent significant learning. LEP children have enough handicaps in school and in life without having them suffer through activities designed to produce spuriously inflated test scores that do not represent true learning.

Situational Factors

Test anxiety. The most pervasive and insidious test score depressant is test anxiety. It has been most extensively measured and researched, and though more research is needed, particularly with LEP children, a strong case in the form of a working hypothesis can be built around this prior research and the assumptions we made about LEP children. In a comprehensive review of test anxiety in the schools, Eccles and Wigfield (1989) submit that text anxiety increases over time and negatively affects school performance. Some factors that seem to contribute to test anxiety are:

1. High stakes tests,
2. Severe time limits on tests,
3. Use of letter grades,
4. Transition from elementary to junior high schools,
5. Poor quality of instruction,
6. Unstructured learning environment, and
7. Negative learning histories.

Given our assumptions about LEP students, the seven conditions cited as contributing to test anxiety seem prevalent in this population. LEP students have more negative learning histories. Negative learning history is also associated with low letter grades, another contributor to test anxiety. Their typically low socioeconomic status creates social conditions by which comparisons with mainstream students leads to lower self-image and lower motivation. If instruction is loosely organized, their test anxiety is heightened. If the learning environment does not fit the culture and the work habits of its LEP students, then the learning environment may serve to increase anxiety. The fact that tests are timed and that LEP students are taking tests in a foreign language must increase their test administration time and reduce their test performance. Besides increasing test anxiety, stress is believed to be a potent factor that also affects test performance (Duran, 1983).

One interesting exception to the above line of reasoning and evidence can be found in a review of American Indian children's test performances by Neely and Shaughnessy (1984). They cite research showing that anxiety is actually lower, so low that it may lead to low test performance.

Timed testing. Some research reports the phenomenon of fast and slow test-taking styles. Knapp (1960) submitted that Mexicans are disadvantaged on timed test because their culture does not promote a fast test-taking style, therefore Mexican children may be disadvantaged in timed tests. The argument and research extends to Native American children. However, as Bridgman (1980) points out, there is very little research to report on the test-taking speed of LEP children.

Examiner effect. Part of test performance can be attributed to the learning environment of the classroom. The role of the examiner on Puerto Rican children was studied by Thomas, Hertzig, Dryman, & Fernandez (1971). They found that performance on an IQ test was increased when the examiner was similar to the child in terms of gender, ethnic background, and fluency in Spanish. Such a study raises an issue that the social context for the test may have some bearing on how hard children try on these tests. Having a teacher who is similar to his or her children may have a positive effect on test performance, and, conversely, differences between teachers and students may have opposite effects.

Setting. Seitz, Abelson, Levine, and Zigler (1975) contend that the site for the test has some effect on children's performances. Their study dealt with disadvantaged children instead of LEP children. However, since LEP children are often disadvantaged, these findings may equally apply to both sub-populations.

Context Factors

Language handicaps. The barrier of learning English and at the same time performing on an achievement test written in that language has to be significant in light of assumptions made earlier about LEP students. As pointed out previously in this paper, huge differences exist between the test scores of LEP and monolingual students in Arizona (Bishop, 1988) and with a national sample (Rock et al., 1991). As one teacher explains (Haas, et al., p. 124):

> Iowa Test of Basic Skills testing regulations discriminate against ESL students. As it takes four to seven years for students to truly become proficient in a second language, especially "academic" language, testing them at grade level after one year on the same level as native speakers is inane.

Fortunately, significant research has been done and is further needed on language proficiency (Duran, 1988). The implication is that before students from diverse educational, ethnic, and social backgrounds can perform on published standardized achievement tests in a mainstream environment, they must first qualify by proving to have a satisfactory level of mastery in the English language. Without such proven proficiency, it would be easy to invalidate test results for LEP children.

Cultural influences. Little research has been reported on the influence of culture on test scores. Nonetheless, there is enough logical and some empirical evidence to suggest that culture plays an enormous role on the success of children. For instance, as previously reported in this paper, in the study by Mine, with others (1987), Japanese parents were shown to negatively influence test anxiety through child-rearing patterns. The study by Knapp (1960), while outdated and about IQ testing, suggests that Hispanic students generally have a different approach to standardized testing. The study by Thomas et al., (1971) shows that the ethnic background and language facility of the examiner may have an influence on test results.

Neely and Shaughnessy (1984) reported that over 300 tribes and 250 languages exist within American Indian culture. These researchers conclude that within this population, and probably other populations, the existence of a different culture is a serious deficit with respect to schooling. For instance, native American children are typically noncompetitive, and do not want to be singled out for recognition. These researchers also point out that most American Indian children speak English only in the schools, therefore the language facility is a serious handicap in a testing situation, because most tests deal with American life that is foreign to tribal children. Such disparities between American Indian children and mainstream children are often cited by teachers as reasons for invalidating standardized achievement test scores (Haas, et al., 1990).

Socioeconomic status. While this fact is obvious to most educators, in evaluation and policy studies, the socioeconomic status of school districts, schools, and children is unnoticed in the reporting of test scores. A considerable relationship exists between family income and test scores (Test Scores and Family Income, 1980). Since LEP children are often of low socioeconomic status, test scores need to be reported in this context so interpretations and uses can be made with the understanding of the handicapping condition presented by low socioeconomic status.

Another factor is *social capital*, a term coined by sociologist James Coleman (1987) that refers to money, other forms of support, and opportunities available to children both inside and outside the home for their growth and development. Coleman believes that social capital is eroding and affecting children's progress in schools. Thus in the interpretation of test scores and the formulation of policy regarding schooling, social capital should be considered as part of the context of the test scores. To fail to consider social capital pollutes test score interpretations and uses.

Summary and Recommendations

1. **Test uses and interpretations should be based on multiple rather than a single indicator.**

   The mindless use of a single score or a set of test scores from a single test is indefensible.

2. **Test results should not be used in ways unintended by its publishers.**

   As indicated in numerous references in this paper, there is gross overreliance, overuse, and misuse of test scores.

3. **Causal interpretations relating to schools and teachers are invalid without considering the full context of causes, and particularly with a test that fails to measure the full scope of school achievement.**

   The need for accountability forces us to make causal attributions about the influences of school on school learning. However, the meaning of any test score, if unpolluted, reflects a lifetime of school and non school learning and a myriad of influences, which partially include, prenatal care, infant stimulation, nutrition, parental support for education, education levels of parents, number of parents in the home, amount of television viewing, degree to which parents read to children, mental ability of parents, economic status, English language facility, developmental status, mental health, family mobility, social capital, motivation, attitude, academic self-confidence, fatalism (locus of control), self-esteem, learning environments in home and school, and quality and quantity of learning in home and school. Many of these factors reside outside of schools.

4. **Interpretations and uses of standardized test scores are often polluted. Extreme caution should be used in interpreting and using test scores for important decisions.**

   We have gained invaluable understanding in the process of aligning curriculum and instruction with testing. The sensible application of this process will lead to better instruction and better outcomes, but all educators and laypersons must understand that outcomes must come fairly and not through deceptive practices such as exemplified in the litany of test score pollution.

5. **We need more wisdom in the definition and measurement of school achievement and sensible, defensible interpretations and uses.**

   As many observers have pointed out, school achievement is not well defined, and therefore its measurement cannot be entirely successful. Also, the general concept of school achievement is changing toward problem solving and other forms of higher level thinking.

6. **Test scores from LEP students appear to be invalid for many interpretations and uses listed in Table 1.**

   While research is woefully inadequate on this topic, enough information exists to suggest that scores obtained from LEP students are going to be very low and language facility blocks both performance and efforts to learn. We need to make certain that test scores are used in ways we can defend and avoid unwise uses of test scores of LEP children.

7. **We need more research to understand the context and motivational factors influencing test performance of LEP students, particularly those students with test anxiety.**

   Sufficient evidence exists to suggest that other factors interfere with the test performance of LEP students. These factors may substantially include motivation.

This paper has identified a problem with the interpretation and use of test scores. The problem has become so serious that standardized achievement tests are being abandoned in favor of "authentic assessment." Unfortunately, the problem is not with the type of test. The problem appears to stem from unwise uses of test results as well as attempt to improve test results through questionable means. The implications for the education of LEP students are significant, because test score pollution may be exacerbated in this context. The recommendations offered here express the concern that the role of testing in instructional programs needs to be more focused around alignment of curriculum, instruction, and tested outcomes. Also, laypersons will need to be better instructed in this role of testing in instructional programs.

Note

[1] A phrase (p. 145) coined by Popham (1987) to describe test results with severe consequences, such as non promotion, the funding of schools or districts, or the awarding of merit pay to teachers or principles on the basis of high test scores.

References

American Psychological Association, American Educational Research Association, National Council on Measurement in Education. (1985). Standards for Educational and Psychological Testing. Washington, DC: American Psychological Association.

Baker, E. L. (1991). Expectations and evidence for alternative assessment. Authentic Assessment: The rhetoric and the reality. Symposium conducted at the annual meeting of the American Educational Research Association, Chicago, IL.

Bangert-Drowns, R. L., Kulik, R. L., & Kulik, C. C. (1983). Effects of coaching programs on achievement test scores. Review of Educational Research, 53, 571-585.

Bennett, R. E., Rock, D. A., & Wang, M. (1990). Equivalence of free-response and multiple-choice items Journal of Educational Measurement, 28(1), 77-92.

Berk, R. A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. Journal of Personnel Evaluation in Education. 1(4), 345-364.

Bishop, C. D. (1988). Statewide report for Arizona pupil achievement testing. Phoenix, AZ: Arizona Department of Education.

Brandt, R. (1989). On misuse of testing: A conversation with George Madaus, Educational Leadership, 46(7), 26-30.

Bridgman, B. (1980). Generality of a "fast" or "slow" test-taking style across a variety of cognitive tasks. Journal of Educational Measurement, 17(3), 211-217.

Cannell, J. J. (1987). Nationally normed elementary achievement testing in America's public schools: How all fifty states are above the national average. Daniels, WV: Friends for Education.

Cannell, J. J. (1989). How public educators cheat on standardized achievement tests. Albuquerque, NM: Friends for Education.

Canner, J. (Chair) (1991). Task Force on Standardized Testing. Washington, DC: National Council on Measurement in Education.

Cohen, S. A. (1987). Instructional alignment: Searching for the magic bullet. Educational Researcher, 16, 16-20.

Coleman, J. S. (1987). Families and schools. Educational Researcher, 16, 32-38.

Dorr-Bremme, D. W. & Herman, J. L. (1986). Assessing student achievement: A profile of classroom practices. CSE Monograph Series in Evaluation, (Number 11). Los Angeles, CA: Center for the Study of Evaluation.

Duran, R. (1983). Hispanics' education and background. New York, NY: College Entrance Examination Board.

Duran, R. (1987). Validity and language skills assessment: Non-English background students. In H. Wainer & H. I. Braun (Eds.) Test validity. Hillsdale, NJ: Erlbaum, pp. 105-127.

Educational Testing Service (1980). Test scores and family income. Princeton, NJ: Author.

FairTest Examiner (1987). 1(1), 1-16.

Foreman, D. I. & Kaplan, J. D. (1986). Scoring high in math. (Subject-centered book B). New York, NY: Random House.

Freeman, D. J., Belli, G. M., Porter, A. C., Floden, R. E., Schmidt, W. H., & Schwille, J. R. (1983). The influence of different styles of textbook use on instructional validity of standardized tests. Journal of Educational Measurement, 20(3), 259-270.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. American Psychologist, 39, 193-202.

Guskey, T. R. & Kifer, E. W. (1990). Ranking school districts on the basis of statewide test results: Is it meaningful or misleading? Educational Measurement: Issues and Practices, 9, 11-16.

Haas, N. S., Haladyna, T. M., & Nolen, S. B. (1990). Standardized testing in Arizona: Interviews and written comments from teachers and administrators. (Technical Report 89-3). Phoenix, AZ: Arizona State University West.

Haertel, E. (1986). The valid use of student performance measures for teacher evaluation. Educational Evaluation and Policy Analysis, 8(1), 45-60.

Haertel, E. & Calfee, R. (1983). School achievement: Thinking about what to test. Journal of Educational Measurement, 20(2), 119-132.

Haladyna, T. M. (in press a). Context dependent item sets. Educational Measurement: Issues and Practices.

Haladyna, T. M. (in press b). The effectiveness of several multiple-choice formats. Applied Measurement in Education.

Haladyna, T. M. (1991). Generic questioning strategies for linking teaching to testing. Educational Technology: Research and Development, 39(1), 73-82.

Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. Educational Researcher, 20(5), 2-7.

Haladyna, T. M., Haas, N. S., Nolen, S. B. (1990). Test score pollution. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Hall, J. L. & Kleine, P. F. (1990). Educator perceptions of achievement test use and abuse: A national survey. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, MA.

Hill, K. & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. The Elementary School Journal, 85, 105-126.

Hoffman, B. (1964). The tyranny of testing. New York, NY: Collier.

Karr, S. K. & Johnson, P. L. (1987). Measuring children's stress: An evaluation of methods. Paper presented at the annual meeting of the National Association of School Psychologists. (ERIC Document Reproduction Service No. ED 285 072)

Koretz, D. M. (1991). The effects of high stakes testing on achievement: Preliminary findings about generalization across tests. In R. L. Linn (Chair), Effects of High-Stakes Educational Testing on Instruction and Achievement. Symposium conducted at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago, IL.

Koretz, D. M. (1989). The new national assessment: What it can and cannot do. NEA Today, 7(6), 32-37.

Koretz, D. M. (1991). State comparisons using NAEP: Large costs, disappointing benefits. Educational Researcher, 20(3), 19-21.

Knapp, R. R. (1960). The effects of time limits on intelligence tests of Mexican and American subjects. Journal of Educational Psychology, 51, 14-20.

Leinhardt, G. & Seewald, A. M. (1981). Overlap: What's tested, what's taught? Journal of Education Measurement, 18(2), 85-96.

Linn, R. L. (1987). Accountability: The comparison of educational systems and the quality of test results. Educational Policy, 1, 181-198.

Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of the claims that "everyone is above average." Educational Measurement: Issues and Practices, 9(3), 5-14.

Madaus, G. F. (1988). The influence of testing on curriculum. In L. N. Tanner (Ed.) Critical issues in curriculum. Eighty-seventh Yearbook of the National Society for the Study of Education (pp. 83-121). Chicago, IL: University of Chicago Press.

Mehrens, W. A. & Kaminski, J. (1989). Methods for improving test scores: Fruitful, fruitless, or fraudulent? Educational Measurement: Issues and Practice, 8, 14-22.

Messick, S. (1984). The psychology of educational measurement. Journal of Educational Measurement, 21, 215-237.

Messick, S. (1987). Assessment in the schools: Purposes and consequences (Research Report RR-87-51). Princeton, NJ: Educational Testing Service. Also appears as a chapter in P. W. Jackson (Ed.) (1988) Educational change: Perspectives on Research and Practice. Berkeley, CA: McCutchan.

Messick, S. (1989). Validity. In R. L. Linn (Ed.) Educational Measurement (3rd ed, 13-103). Washington, DC: American Council on Education.

Mine, H., with others. (1987). A study of the effects of child rearing patterns on test anxiety in late adolescents. Paper presented at the biennial meeting of the International Society for the Study of Behavioral Development. Tokyo, Japan. (ERIC Document Reproduction Service No. ED 289 634)

Neely, R. & Shaughnessy, M. F. (1984). Assessments and the Native American. (ERIC Document Reproduction Service No. ED 273 889).

Nickerson, R. S. (1989). New directions in educational assessment. Educational Researcher, 18, 3-7.

Nitko, A. J. (1989) Integrating teaching and testing. In R. L. Linn (Ed.) Educational Measurement (3rd ed.). Washington, DC: American Council on Education.

Noggle, N. L. (1988). Report on the match of standardized tests to Arizona essential skills. Tempe, AZ: College of Education, Arizona State University, School Personnel Evaluation and Learning Laboratory.

Nolen, S. B., Haladyna, T. M., Haas, N. S. (In press). A survey of actual and perceived uses, test preparation activities, and effects of standardized achievement tests. Educational Measurement: Issues and Practices.

Nolet, V. & Tindal, G. (1990). Evidence of construct validity in published achievement tests. Paper presented at the annual meeting of the American Educational Research Association, Boston MA.

Paris, S., Lawton, T. A., Turner, J. C. & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. Educational Researcher, 20(5), 12-20, 40.

Paris, S., Turner, J. C., & Lawton, T. A. (1990). Students views of standardized achievement tests. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Plass, J. A. & Hill, K. T. (1986). Children's achievement strategies and test performance: The role of time pressure, evaluation anxiety, and sex. Developmental Psychology, 22, 31-36.

Phillips, S. E. & Mehrens, W. A. (1987). Curricular differences and unidimensionality of achievement test data: An exploratory analysis. Journal of Educational Measurement, 24(1), 1-16.

Popham, W. J. (1987). The merits of measurement driven instruction. Phi Deltan Kappan, 68, 679-682.

Popham, W. J. (1990, January). Appropriateness of teachers' test preparation practices. Paper presented at a Forum for Dialogue Between Educational Policymakers and Educational Researchers. UCLA Graduate School of Education and the California School Boards Association, University of California, Los Angeles, CA.

Rock, D. A., Pollack, J. M., & Hafner, A. (1991). The tested achievement of the National Educational Longitudinal Study of 1988 Eighth Grade Class. Washington, DC: US Department of Education.

Sarnacki, R. E. (1979). An examination of testwiseness in the cognitive test domain. Review of Educational Research, 21, 252-279.

Seitz, V., Abelson, W. D., Levine, E., & Zigler, E. (1975). Effects of place of testing on the Peabody Picture Vocabulary Test scores of disadvantaged Head Start and non-Head Start children. Child Development, 46, 481-486.

Shepard, L. A. (1989). Why we need better assessments. Educational Leadership, 46(7), 4-9.

Smith, M. L. (1990). The meanings of test preparation. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. Educational Researcher, 20(5), 8-11.

Smith, M. L. & Rottenberg, C. (In press). Unintended consequences of external testing in elementary schools. Educational Measurement: Issues and Practices.

Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C., and Cherland, M. (1989). The role of testing in elementary schools. Los Angeles, CA: University of California at Los Angeles, Center for Research on Educational Standards and Student Tests, Graduate School of Education.

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. Journal of Educational Measurement, 26, 233-246.

Thomas, A., Hertzig, M., Dryman, I., Fernandez, P. (1971). Examiner effect in IQ testing of Puerto Rican working-class children. American Journal of Orthopsychiat, 41(5), 809-821.

Toch, T. (1991). In the name of excellence: The struggle to reform the nation's schools, why its failing, and what should be done. New York, NY: Oxford Press.

Walberg, H. J. (1980). A psychological theory of educational productivity. In F. H. Farley & N. Gordon (Eds.) Perspectives on educational psychology. Chicago, IL & Berkeley, CA: National Society for the Study of Education & McCutchan Publishing.

Wardrop, J. L., Anderson, T. H., Hively, W., Hastings, C. N., Anderson, R. I., Muller, K. E. (1982). A framework for analyzing the inference structure of educational achievement tests. Journal of Educational Measurement, 19(1), 1-18.

Wodtke, K. H., Harper, F., Schommer, M. & Brunellia, P. (1990). How standardized is school testing? An exploratory study of standardized group testing in kindergarten. Educational Evaluation and Policy Analysis, 11(3), 223-236.

Wright, B. D. & Stone, M. (1979). Best test design. Chicago, IL: University of Chicago MESA Press.

Zatz, S. & Chassin, L. (1985). Cognition of test-anxious children under naturalistic test-taking conditions. Journal of Consulting and Clinical Psychology, 53, 393-401.

Response to Thomas Haladyna's Presentation

Gary Hargett
University of New Mexico

My discussion of Dr. Thomas Halaydna's paper will be in two parts. First, I want to talk about a kind of test score pollution that Dr. Thomas Halaydna mentioned which has to do with the public perception of test scores and put my discussion in the context of the debate on educational reform. Secondly, I will offer some thoughts that are more specific to implications of testing for LEP students in the context of Title VII program evaluation which is originally what I was asked to do for this symposium anyway.

Dr. Thomas Halaydna and his paper listed several sources of test score pollution that invalidate test scores. The sources include tailoring curricular to specific tasks, coaching students for tests, teaching test wiseness, even excluding low achieving students from taking standardized tests. All of these we know happen, but I would suggest that maybe the greatest source of test score pollution behind these other sources is one that he has alluded to and I think needs further attention and that is the disproportional importance attached to tests by policy makers, editorialists and other commentators based on misconceptions about the role of tests and misinterpretations of the meaning of test scores.

I can illustrate this with a recent example. As you know, just last week the SAT scores were released. This year students in Oregon where I live had the highest average among the states, which got favorable local press. But I was listening to an editorial on television and the commentator thought it was really shameful that our average was only somewhat over 400 on the SAT which he said was only fifty-some percent of the maximum possible score of 800. He clearly does not have any concept of what SAT scores are -- about standard scores and that kind of thing. And I wondered if he really thinks that students take the SAT, sit there and answer 800 questions on each section of the test. But I think that his misunderstanding exemplifies the thinking of many of the loudest critics of the schools who just don't know what normal test scores are all about, and I have even seen this type of misunderstanding at the level of school superintendents who really should know better.

I think like most members of the public, policy makers and public commentators want information on how much this nation's students know, whether they are achieving at grade level, which is itself a not very well-motivated construct; and in fact standardized test scores that are most commonly reported, stanines, percentiles, grade equivalent scores, and NCEs, really don't tell how much a student knows about the subject or what specifically a student knows even if we can assume a high degree of content and construct validity, which we know is not a safe assumption. The only kind of information these scores really convey is the degree to which a student is at, above, or below the average of the test norming group. These tests were developed on the assumption that I was taught on my own measurement training which is that we are interested in differences among people and we want to obtain reliable measurements of true differences among people. We want to know who's best, who's worst and who's in between.

These scores can be construed to represent content only if we share a common concept of content at each grade level. Yet, Dr. Thomas Halaydna gave the example of his own state, Arizona, where they found that the content of the test that had been used did not correspond to what the state curriculum was mandating be taught, and I think most of us have had this kind of experience; even among curriculum text publishers you have a very large variation of what a publisher construes to be grade level kind of work. So the very logic of norm referenced standardized tests may be inconsistent with the kinds of interpretations that most policy makers and editorialists try to impose upon them. These people want to know what our students know. The tests really only tell who knows more and who knows less. What worries me is the implications of the continued use of these kinds of tests in educational reform where there is emphasis on competition and, in my opinion, not a very healthy emphasis.

Any norm referenced test free of score pollution will find half the students above average and half below average. Now we are facing proposals for universal national testing and in an atmosphere of academic competition I think there will be a lot of hand wringing over who is below average and laying a lot of blame usually at the steps of the school. I think this will be true even if subsequent generations of students do achieve more than their predecessors. This approach to testing does not promote the principle of excellence for all. It only invites comparisons which some policy makers do want and which norm referenced tests can provide. But it still doesn't really tell what our students have learned.

I think there are plenty of examples of how test scores are perceived. The August 25 issue of Parade Magazine, which as you know enters millions of homes every Sunday, headlined an article about the schools with a statement about declining test scores. Yet we know, and as Dr. Thomas Halaydna mentioned a few minutes ago, achievement test scores have not declined, they have risen, and presumably for reasons due to test score pollution as he pointed out. It has not been suggested to my knowledge that in some cases the scores may have risen because schools are really doing a good job. The current orthodoxy and, in fact, it's almost a national policy now, is that our public schools are a failure and there aren't any real criteria for that judgment, and it's contrary to much of the objective evidence that does exist. Dr. Thomas Halaydna referred to the Cannell report that came out a few years ago and that was discussed in the Fall 1990 issue of Educational Measurement: Issues and Practice. In that issue, I found it interesting that many explanations were offered as to why standardized achievement test scores were inflated, due mostly to pollution, but I still did not see any evidence for the decline of American education. Laurie Shepard's article in that issue cited data from the National Assessment of Educational Progress that showed modest gains and cited findings from the congressional budget office with figures that also showed improved achievement, just not as dramatic as the gains that are shown on the standardized achievement tests.

I don't mean to suggest that we are not facing real serious educational problems and I certainly don't suggest that we should not be seriously discussing educational reform. I think, of course, we can do better -- we should be doing better. But I think we should take a hard look at our expectations for student achievement and I don't think we should base our discussions on the a priori premise that the schools have failed without any solid evidence to that effect. The evidence as far as I can tell is pretty much anecdotal. I'd like to give an example of my own state of Oregon which has recently been nationally praised for taking the lead in educational reform. You may have heard about our reform package that was passed by the legislature just this summer. I think the point of view of most Oregon educators is that our legislators enacted a reform package without any clear statement of what the problems were or any compelling linkage of the reforms to those problems.

At the Seattle hearings on the national goals, I heard Dr. Ramsey Seldon, who is a member of the National Goals Panel Resource Group, remark that at this point we really don't know what's going on in the schools. He says, for example, we don't even know how many teachers and how many schools are using skills based as opposed to whole language reading approaches and to what degree they are using them. In other words, we're clamoring for reform without necessarily knowing what it is we are trying to change.

Dr. Thomas Halaydna's conclusions about the problems of LEP students taking standardized tests are certainly valid and they point up certain problems associated with recent proposals for universal testing. I refer to the proposal that every student should take an achievement test or a series of such tests at certain points in his or her educational career. I think we have to look at the implications of this kind of universal testing and I would suggest that we do not need universal testing to assess the attainment of educational goals assuming, that is, that tests can be calibrated to those goals or if tests are the most desirable measure of goal attainment.

I think we can accomplish that through well-applied matrix sampling, which is what the California Assessment Program does. The only reason for obtaining test scores on every individual is if there are individual consequences and implications based on the individual's test score. I recently heard a spokesman for a group called Educate America, advocate testing of all high school students in the fall of their senior year. In his comments he said that at first this would be low stakes testing but, then, when it was pointed out that students are not motivated to do well on low stakes test, he said students will be motivated to do well because these scores might be considered in college admissions or looked at by potential employers. Well, at this point these become high stakes test scores.

I agree with the observation that, for many or most purposes, test scores for LEP students tend to be invalid. I think they're valid in one sense, in the logic of norm referenced tasks that LEP students don't know as much or don't have the same kind of skills as the norming group on whatever it is the standardized tests measure. Whether that's important or whether LEP students have academic talents that are not measured by the tests is a separate issue.

I personally do not advocate large scale high stakes testing, but I am worried about certain implications of the exclusion of LEP students from such tests even out of benign concern for the invalidity of their test scores. My most important concern is that this sends a message that marginalizes LEP students, that since we cannot test them they're marginal to education. If a point of tests is to drive excellence in education, they should drive excellence for LEP students as well. My other concern is that scores from large scale high stakes test may become another kind of credential. Rightly or wrongly, the high school diploma is widely perceived as not necessarily representing the mastery of academic skills. That is part of the reason for the demand for new tests such as the minimum competency tests we have seen in many states. If LEP students are excused from tests because their test scores are invalid due to language, they will be leaving school without an important credential.

I think we are seeing the possibility of this in Oregon where part of our reform package is that, at tenth grade, students will take a test for a Certificate of Initial Mastery -- whatever that means. And after that, they go into either a college prep track or a vocational track, which has many of us sort of in horror. But if the LEP students cannot take these tests for the Certificate of Initial Mastery, then you wonder, well, what options are open to them after the tenth grade? I don't mean to imply that I favor testing LEP students with tests based on English only norms because I certainly don't. I don't even favor developing alternative norms because I think that would be pointless and probably impossible. I am only pointing out some logical consequences for LEP students in the context of large scale testing. My personal preference is that we back away from the imposition of high stakes testing for all students.

This brings me to the second part of my discussion dealing with the use of tests, particularly standardized achievement tests and Title VII program evaluation. The Title VII regs do not require standardized tests. They require reports of educational progress measured as appropriate by tests of academic achievement and they require that the evaluation instruments that are used consistently and accurately measure progress toward the project objectives, that they be appropriate considering several factors including language proficiency, and that they be administered at twelve month testing intervals. I think that many people have construed this to mean that standardized tests are required because of the key terms "academic achievement" and "twelve month intervals". They may also think that since the tests they use have to be reliable and valid, they should use standardized tests because, after all, these have technical manuals that report their validity and reliability.

However, as Dr. Thomas Halaydna has pointed out, these are not reliable and valid tests for LEP students for a number of reasons, including lack of content validity for a typical Title VII project curriculum. What they most reliably do is show that LEP students perform much lower than other students measured by these tests, which is not surprising since part of the definition of LEP is that they are not able to learn successfully in classrooms for the language of instruction and the testing is in English. What Title VII evaluation and regulations call for is a measure of progress toward accomplishing the objectives of the project. It's not uncommon to see Title VII project objectives written in terms of bringing the LEP student up to grade level.

But I think we need to think about the implications of this kind of project objective and how to test it. It would seem on the face of it that standardized tests would be a logical measure of that kind of objective. But I see two problems apart from the obvious question of what grade level even means. We lose sight of the fact that grade level is not a point but a range of abilities. The first problem is whether this kind of objective is reasonable for many projects, especially if you consider projects that are serving some older students -- upper elementary and high school students who may be coming into the schools with very weak academic preparation in their own native languages. It's probably not reasonable to expect them to perform comparably to the norm group on the standardized achievement test or in other measures as well.

Where I have seen these tests most effectively used has been with projects that work with early elementary students and are able to give sustained service over a period of years and, in fact, a service that is actually mainstreaming from the very beginning. It's not the model where first we give them Title VII and then we give them the real curriculum. I think that any bilingual education program should strive to help the students advance as much as possible in language and academic abilities, but if the measure of gain is performance on a standardized achievement test and the goal performance is comparable to the norming group, that may be an elusive goal. I would like to see Title VII projects experiment with some of the alternative assessment approaches that are being discussed in this symposium. One reason for this is something I've learned during my experience with program evaluation both through the EAC-WEST and other evaluation roles I've played. I've learned that evaluation issues become a focal point, maybe even a lightning rod, for the discussion and clarification of many other issues.

We've seen this in the national debate on education. Unfortunately, this debate is murky because the evaluation issues are not well understood. But I think there is a great potential for the role of performance or authentic assessments in Title VII evaluation. I think first of all that many, maybe most Title VII project curricula, really are not built around the kind of things standardized tests are intended to tap into. Therefore, the projects need assessments that are built around the curricula, and we hope that those curricula are targeting levels of excellence and meaningful tasks and applications. I think that the development of performance assessments provides the form for articulating expectations, thereby setting standards of excellence to teach toward. I think that's a more exciting educational concept than either grade level or minimum competency. By the way, this is not an easy process, as the people who have been working on performance assessment can tell you. From my own experience in many of the workshops I have given, one of the hardest things to do is to get teachers to articulate the outcomes they expect for their students, and this is true of many kinds of teachers, not just teachers in Title VII programs.

But this is what teachers and other educators have to do in order to meet the kinds of standards of excellence that AMERICA 2000 is supposed to be about. I'm afraid, that if educators don't articulate the expectations, then politicians will, and I personally have more confidence in the educators than the politicians to do a good job of that. Developing performance assessments can have several advantages because by their very nature they set standards of excellence, and I think that's an attitude that Title VII programs must assume, and move away from the deficit model. We know that all students, including LEP students, tend to meet expectations, so we should have expectations that embody excellence. Other potential advantages of performance assessments are high curricular validity and the communicability of test findings. As I suggested earlier, there's nothing really wrong with standardized test scores, but the way they are interpreted miscommunicates the content of the scores, whereas performance assessments are couched in terms of actual performance, what students really can do and how well they communicate.

I don't want to give the impression that performance assessment is an automatic panacea. For one thing, it is a supplement to, not a replacement for, other kinds of assessment. They still do have their place and they do have pitfalls which Dr. Eva Baker went into yesterday. By the way, I think the biggest pitfall is trying to impose performance assessments in the traditional setting. I think the performance assessment only makes sense in an atmosphere where students are performing, and problem solving is part of their everyday educational experience. So if you do want to develop performance assessments for your Title VII project evaluations, I would encourage you to do so but look for guidance and, of course, the EACs are a good place to start looking for that guidance.

To summarize my remarks, I agree with Dr. Thomas Halaydna that test scores have become polluted. The proliferation of test scores might itself be said to be polluting, but the most dangerous pollution is the over-interpretation and the misinterpretation of test scores which I think leads to many of the other sources of pollution that Dr. Thomas Halaydna listed. And I also think that the standardized test should be used cautiously with Title VII evaluations and the Title VII project should consider alternative methods of assessment that promote excellence.

Response to Thomas Haladyna's Presentation

María Pennock-Román
Educational Testing Service, Princeton

Overall, I concur with Haladyna on many points, and I agree with most of his recommendations about proper and improper uses of tests. Nevertheless, I find that his application of the labels "pollution" and "contaminants" obscures the issues at hand, beginning with the title. If one looks closely at most of Haladyna's criticisms, it is evident that he disapproves of common uses of tests by state policy makers, school administrators, and teachers. For this reason, I believe it would be more appropriate that his paper be titled "Test Use Pollution."

My reaction to his major points are summarized in three tables in order to conserve space and time. Most of the entries in the tables are self-explanatory so that only selected rows will be discussed.

Desirable and Undesirable Test Practices

Table 1 presents a contrast between Haladyna's opinions and mine concerning what testing practices are desirable or inappropriate. Next to each testing practice that Haladyna considers a "contaminant" in test scores is his classification as to whether the practice is ethical (E) or unethical (UE). In the adjacent column are my views concerning this classification and comments to explain my rationale.

As shown in Table 1, the author in some ways contradicts himself as he applies the negative label of "contaminants" to testing practices that he himself considers "ethical." Haladyna on the one hand considers training test wiseness or increasing student motivation as contaminants of test scores but, on the other, he classifies training in test wiseness and increasing motivation as ethical practices. Later, he makes a recommendation that LEP students be trained to take tests properly.

Happily, there is a fairly easy way to resolve this inconsistency by changing the form in which the "contaminant" is described. For example, I believe that in the case of students outside the mainstream it is inexperience with tests or test naïvete that may add unnecessary noise to scores. In a study of test-taking skills of Hispanic junior and high school students in California (Pennock-Román, Powers, & Perez, 1991), I was appalled to find that even filling out answer sheets presented problems for some students. Certainly, test naïvete may reduce the validity of the test for inexperienced test tak-

Table 1
Juxtaposition of Haladyna's and Pennock-Roman's Views of Testing Practices

Haladyna's Views		Pennock-Roman's Views		
Testing Practices Considered "Contaminants"	Ethical/ Nonethical	Ethical/ Unethic.	Contaminant?	Rationale
1) Preparing students for tests				
a) Training in testwiseness	E	E	No	Extreme test naivete may introduce irrelevant variance. See Maspons & Llabre (1985).
b) Checking answer sheets for proper completion	E	E	No	See above
c) Presenting items similar to those on the test	UE	E	No	See above
d) Using test preparation books such as Scoring High	UE	E/UE	Some Cases	Depends on test preparation book. Good if reviews fundamental math, etc.
e) Presenting items verbatim from the test	UE	UE	Yes	Introduces variability in scores irrelevant to domain of achievement.
f) Cheating (giving out answers or hints, correcting answer sheets)	UE	UE	Yes	Introduces variability in scores irrelevant to domain of achievement.
2) Promoting student motivation for the test	E	E	No	It is lack of motivation that is irrelevant to the measurement of academic achievement.
3) Teaching to the test				
a) Preparing objectives based on items on the test and teaching accordingly	UE	E/UE	Usually	Depends on completeness of test. In some instances, (e.g., Advanced Placement Tests), objectives are comprehensive.
b) Developing a curriculum to match the test	UE	E/UE	Usually	Same as above.
4) Dismissing low-achieving students on testing day to artificially boost test scores.	UE UE	E/UE	Some Cases	Occasionally justified, but criteria for exclusion should be uniform and well defined.

ers. A small-scale study by Maspons and Llabre (1985) lends support to this view, but a lot more research needs to be done in this area.

He and I also agree in our disapproval of adapting curricula and "teaching to the test," under most circumstances. However, I can think of exceptional cases where especially comprehensive tests can serve as good guides to curricula. At the risk of sounding as though I'm putting in a "plug" for my company, consider the Advanced Placement (AP) Tests which are college-level achievement tests in various subjects. Students attaining high grades on a given Advanced Placement Test receive college credit for that course. These tests have been carefully designed to cover a domain area quite rigorously and thoroughly under the guidance of college professors from representative universities. Because of the care in its construction, curricula designed to encompass the material of an AP test may indeed be a good one to follow.

However, tests such as the AP tests are the exception rather than the rule. In general, "teaching to the test" is not a good idea because most achievement tests are not linked to specific, well-defined courses of study.

Haladyna and I also concur in disapproving of the practice of dismissing low-achieving students on testing day to artificially boost test scores. One exception mentioned later on by Haladyna are LEP students who should be excused from standardized achievement tests until their competency in English is sufficiently high to make test scores meaningful. Of course, defining the point at which there is enough proficiency in the language of the test is a difficult task. More research is needed in this area. Besides LEP students, there are other groups of exceptional children who are learning disabled or physically handicapped for whom traditional tests may be invalid. These students ought to be excluded from analyses of summary statistics for a given school.

In any case, the criteria for exclusion of special children from public reports of test results need to be well-defined. Results will be comparable across school districts only when such criteria are applied consistently on the districts that are contrasted. It would be desirable if such criteria could be defined on a national basis to make norms on widely used achievement tests more useful.

Table 2
Juxaposition of Haladyna's and Pennock-Roman's Views Concerning Relevance of Variables to Domain Tested

Haladyna's Views	Pennock-Roman's Views	
Variables Considered "Contaminants"	Contaminant?	Rationale
1) Situational Factors		
Test Anxiety	Sometimes	Excessive anxiety or nonchalant attitude are problems.
Stress	Yes	
Speededness of the test	Usually	But in rare cases, speededness part of construct.
Examiner Effects	Sometimes	Probably little effect with multiple-choice tests
2) Context		
Language deficits	Usually	Influence and relevance of language deficit varies by test content. Quantitative tests less of a problem.
Socioeconomic context	No	For achievement tests, these sources are related to underlying content area.
Family Mobility	No	See above.
Family and Home Influences	No	See above.
Prenatal/Early Infant Influences	No	See above.

Contextual and Situational Factors

As shown in Table 2, Haladyna and I are also largely in agreement with regard to the issue of speediness and language deficits as a contaminant in standardized test scores. He should consider adding to his paper some recent reviews of the literature on speediness for non-native speakers of English which support this point of view (Llabre, 1991; Pennock-Román, in press). There is evidence from many sources, that non-native speakers of English have great difficulty in completing selective-admissions tests, particularly the verbal portions of tests such as the SAT, GRE, and GMAT (see review by Pennock-Román, in pr[illegible])

In Pennock-Román (1990) and in the aforementioned review (Pennock-Román, in press), there is also a discussion of language proficiency in the language of the test as a factor that interferes with the measurement of ability and achievement. However, one finding of special interest is that some curriculum-specific achievement in subject areas are somewhat less influenced by language proficiency than more global types of ability tests. Naturally, quantitative tests are less influenced by language factors, but more verbal types of tests show this effect also. One explanation is that non-native speakers of English may be on more equal footing with mainstream students in regard to academic vocabulary (e.g., technical terms in science) than they are with language terms learned mostly outside of the school environment (e.g., names of fruits, furniture).

In contrast, Haladyna and I differ in our positions concerning family background and other contextual influences on test performance. He is somewhat ambivalent in this position concerning the classification of these variables as contaminants or meaningful variance. Whereas, he lists socioeconomic context, family mobility, family and home influences as "documented sources of test score pollution," on page 31 he states that "Any test score, if unpolluted, reflects a lifetime of school and non school learning and a myriad of influences." Hence, his position is not clear -- are home influences pollution or not?

From my perspective, background factors affect the quality of training a student has had, which for the most part is a valid source of variance because it does affect the content domain to be measured (academic achievement) in our society where educational resources are unevenly distributed. On the other hand, these sources do limit the uses that test scores can serve. And it is not proper that teachers and schools should be evaluated without taking into consideration these factors. Thus, there are problems with using student test performance to evaluate teacher effectiveness because teachers are only one of many influences on those scores. Multiple indicators are necessary to evaluate schools and teachers. I believe that we need to make a distinction here in terms of the different uses of tests and relevance of these variables to the purpose for which the test serves.

Construct Validity Should Refer Only to Intended and Recommended Uses of Tests, Not to All Other Uses That Occur

While the Haladyna uses the APA Standards definition for what is proper evidence of validity, he states that "construct validation calls for the collecting of evidence to support any of the 29 different uses [referred to in his Table 1]," whether it is recommended or not. This is clearly not the intent of the standards. Key words in the Standards are "Evidence ...presented for the major types of inferences FOR WHICH THE USE OF A TEST IS RECOMMENDED... Support the particular mix of evidence presented for INTENDED uses."

He implies that performance tests, alternative testing, and "authentic" measures will provide a future solution, because they are free of the problems that multiple-choice tests have. However, as he points out, the main problems stem from misuse and misapplication of multiple-choice tests. Won't future, performance and alternative tests be subject to misuse also? And, given the many problems in scoring such tests because of subjectivity in grading, won't the potential for misuse be even greater?

As long as we continue to blame the test rather than school and state policies for improper test use, problems will not be corrected, and they will recur with any kind of test that is devised, standardized or not.

His Recommendations Are Mostly Points of Agreement between Us

My points of agreement or disagreement on recommendations are presented in Table 3; you can see that there are few disagreements with the recommendations, and most are self-explanatory. I'd like to suggest that he repeat in the latter pages (pp. 30-31) some of the points referred to earlier in the manuscript, because many are worth reiterating.

Table 3
Evaluation of Haladyna's Recommendations

Pennock-Roman's View of Haladyna's Recommendations	Haladyna's Recommendations
1) Testing Practices	
Mostly Agree	"All children should have comprehensive test-taking courses and be equally skilled in test-taking." (p. 24)
Agree	Should avoid test activities designed to produce spuriously inflated test scores that do not represent true learning. (p. 25)
2) Situational Factors	
Agree	Test anxiety may be reduced through ethical test preparation activities. (p. 25)
3) Context	
Mostly agree	One must first prove that students have a satisfactory level of mastery in the English language before being tested in English. (pp. 28-29)
Mostly agree	Scores should be reported by SES. (p. 19)
Agree	One should take into consideration students' SES and other factors when evaluating teachers and schools using scores. (p. 31)
Summary Recommendations	
Agree	"We should resist attempts to interpret or use test results in ways unintended and unsupported by validating evidence." (p.5)
Agree but issue not well developed in his paper.	"Need more wisdom in the definition and measurement of school achievement." (p. 31)
Agree	"Test scores obtained from LEP students appear to be invalid for many [typical] interpretations and uses." (p. 31)
Agree	"Use multiple indicators for evaluating schools and teachers." (p. 30)
Disagree	"Standardized test scores are often polluted. We should avoid using such test results." (p. 31)

BEST COPY AVAILABLE

Conclusion

In general, most of the criticisms and recommendations that Haladyna makes are sensible; many have been suggested before by measurement specialists and other educators, so there is relatively little new here. The majority of his criticisms do not address test content, format, or test construction. However, by using the term "test score pollution" he puts the blame for many wrong uses of tests on the instruments themselves, rather than on test users. Furthermore, there are some contradictions introduced by grouping too many things under the label of pollutants.

Problems with the use of the terms "pollution" and "contaminants" arise for two reasons. First, these terms, which are loaded with negative connotations, are applied in an overinclusive manner to a variety of practices considered both ethical and unethical according to Haladyna himself. Hence, the label of "contaminants" tends to obscure his distinctions between appropriate and inappropriate uses of tests, thus making his policy recommendations unclear. Second, I find that, in this controversial area, the use of inflammatory language is counterproductive. It interferes with the constructive dialogue among test specialists, educators, and advocates of LEP children that is necessary for positive solutions to measurement problems.

References

Llabre, M. M. (1991). Time as a factor in the cognitive test performance of Latino college students. In Deneen, J., Keller, G., & Magallan, R. (Eds), pp. 95-104, Assessment and access: Hispanics in higher education, NY: SUNY Press.

Maspons, M. M., & Llabre, M. M. (1985). The influence of training Hispanics in test taking on the psychometric properties of a test. Journal for Research in Mathematics Education, 16, 177-183.

Pennock-Román, M. (1990). Test validity and language background: A study of Hispanic American students at six universities. New York: College Entrance Examinations Board.

Pennock-Román, M. (in press). Interpreting test performance in selective admissions for Hispanic students. In Geisinger, Kurt (Ed.) Psychological Testing of Hispanics, Washington, D.C.: American Psychological Association.

Pennock-Román, M., Powers, D. E., & Perez, M. (1991). A preliminary evaluation of Testskills: A kit to prepare Hispanic students for the PSAT/NMSQT. In Deneen, J., Keller, G,., & Magallan, R. (Eds)., pp. 243-264, Assessment and access: Hispanics in higher education, NY: SUNY Press.

LEA Title VII Program Evaluations: Panel Presentations

Raj Balu
Chicago Public Schools

I'm going to present my point of view from the perspective of an administrator. I was an evaluator, but I have been removed from that for the past five or six years. I've been an administrator of bilingual programs in the Chicago public schools a few years. First, I will be presenting information regarding the bilingual education program in the Chicago public schools. Then, I will move on to Title VII programs specifically and what kind of evaluation we are doing. From there, I will talk about the problems we have faced in evaluating the program and some strategies and recommendations. That's the basic outline of my presentation.

Chicago has about 45,000 LEP students, and the state mandates that bilingual education be provided for every child who has been identified as a LEP student. As part of meeting the mandates of the requirements, Chicago public schools assess every child who enters the school system, including the English monolingual and English background children, using a home language survey as to their language background. If any student is identified as coming from a home where a language other than English is spoken, then that child is further assessed in terms of his or her English language proficiency.

Each school has a computer terminal and the schools -- the staff in the schools -- enter the information online, and this information is available online in a central computer system. At the time the data are collected, the student is categorized as knowing no English at all, a little bit of English, or a lot of English but still needs some assistance, or is capable of functioning in a classroom where English is the only language of instruction. For any of the first three, children are categorized as limited English proficient children, and they are provided with bilingual education programs. At this time, we have about 320 schools providing bilingual education programs in about 74 different languages and about 1500 state certified bilingual teachers. Bilingual or ESL-certified teachers provide services to these children. A total of about $36 million is spent by the local schools, about $28 million by the state, and about $2.5 million from the federal government's different programs. That is the range of the program's expenditures.

I have a handout that provides these kinds of statistics, but it is limited in quantity. If you don't get the handout, please provide your name and we'll mail it later. So the online data keeps track of the student's progress, and annually, at the end of the year, during the spring -- March, April, May -- each student is assessed using a citywide program. Those data are also sent to the school and, during the summer, the child's English language proficiency is reassessed, and the student is recategorized if he or she shows additional English language proficiency and is ready to move out of the program. Those not ready to move on stay in the program. This is in terms of the city of Chicago, following the state mandate. As you might have heard before, Chicago is going through the school reform initiative that was passed through the state legislature, that is, each school has a local school council, and the local school council has the right to implement the program that it wants implemented in its schools. Of course, mandated programs, such as bilingual programs, have to be provided in that school.

We have a research evaluation department, and we have a language and cultural education department. The difference between the two is that research and evaluation is in charge of managing the data, providing the evaluation reports and research reports, and planning current studies. The language and cultural education department administers the program and provides technical support to teachers in planning and implementing the programs. The two work in tandem serving the bilingual education program and providing an annual evaluation report to the state within about a six-month period after the program period. Last year we had two-fifth year programs, Title VII programs, and two new programs. One of the new programs is for Arabic bilingual education, and the other one is a developmental bilingual education program.

The developmental bilingual education program has about six schools with specific programs. All of them are Spanish-English developmental programs, and they started functioning as a program during the middle of last year, starting with kindergarten grade level. The plan is to follow the children from kindergarten through the eighth grade, even after the Title VII funding ceases. When we received funding for the developmental bilingual program, we came for a management institute, and a packet of evaluation forms and data collection forms was given to us. Those sets of forms were very useful in planning the evaluation actively for those programs. Later I will explain why they were helpful.

In the Arabic bilingual program, what we have now is that same data that we collected for the state bilingual program, the online data. Annually we are required to provide evaluation reports to the federal government for Title VII bilingual programs. I am sorry to say, we have been consistently late. I don't know whether anyone else has been late, but we have been late, and there are reasons for that. At the same time, I can assure you that the developmental bilingual program evaluations report will not be late, because there was a system set up for that. It asks what things are not being done now. What are the reasons why we are not doing things that we are supposed to do? Or, what are the reasons why we would like to improve upon what we are doing in other departments?

Just within the last two months, the research evaluation department sent home 50 of its 75 staff members. The reason? Budgetary reduction, administrative staff cuts, and a perception in the schools that the administrative structure is too big. Out of the 50 people who were sent home, three were bilingual evaluators, paid out of state bilingual monies. The system is not saving any money, but in order to reduce the perception that the bureaucracy is smaller, the three were sent out. What will happen to the evaluation report that is due within the next year? It is going to be a little bit late.

Another problem we are facing is that local commitment for evaluation and research is not as big as it was in the past. Now the commitment is toward the community to improve the school's program and to show that the test grade equivalent score, even now that's being used in Chicago, has improved a lot, and the students have reached the national norm as required by the state legislature. And as the commitment is for doing the citywide testing at the end of March, or April, or May, and show that the test score has improved. Further, this assessment is required for the city regarding the general program of instruction, not just bilingual education. All of our students are part of that program, part of that assessment. So what happens? The priority for writing that report goes to the citywide testing data report and bilingual education's report stays back.

The third problem or concern in having an evaluation done is about local autonomy, the local school council I mentioned earlier. The council has the autonomy to decide what test it administers, how it administers, how it uses the data, and how it reports the data. Still, we maintain a little control of that because of the state mandate, otherwise we would have lost that control also. I'm not saying that we should control but, to collect the data that are uniform and usable citywide, control is needed.

Under the change in administration the principals are under the local school councils. They are hired/fired by them. The councils are for two years, the principals' contracts are for three years, and when a new council comes the old principal is fired and a lot of principal changes are happening. Increased costs for consultants for evaluating now average $200 per day in Chicago and non-consistent availability of external evaluators and writers. We need writers, but at the same time we hired the students from the local university to write the evaluation report, to do the evaluation, but they are not available all the time.

What we are doing now to overcome these problems and to improve upon the system is to pool the resources. Most of the Title VII evaluation that we are now doing uses part of the state money also and, as such, the state evaluation supplements the evaluation of the Title VII program. We are using college and undergraduate students as testers, and they are sent to schools to provide additionally needed testing. We are assigning staff members to be in charge of the programs, those who have an interest in a particular program. For example, our developmental program manager, there is no specific manager funded by the program, but we have assigned somebody to be in charge of the program, in addition to other duties, who has a personal interest in the program and will do a better evaluating task. In terms of Title VII guidelines, I am going to discuss the three major categories.

We collect student data diligently. We have it online, and it is available for us to analyze and study. Technical standards are maintained in terms of selection, administration, and training. We do not collect as much implementation data as we would like to collect because it is staff intensive. In the developmental program package that was given to us, we have shared that responsibility among the staff of the program in the schools because there are specific forms to be filled out by teachers and the principal of the school with regard to the data that is needed and that is supposed to be collected. In general, we collect implementation data and run it through the state bilingual programs' compliance review, which is done for a third of the schools. So a third of the schools get it without a problem. Other schools have a problem.

Lastly, I would like to talk about a few recommendations. Just like the developmental program package that contained specific data collection blank forms, which is similar to the abstract that was presented earlier by Tomi, it's a little bit more lengthy in detail, and it is divided up into different people's responsibilities to fill out those forms. This would be good for all programs. In order to collect implementation data, it may be good to have support from the evaluation assessment centers instead of the local school collecting those data. If possible, evaluation assessment centers have funds to have local persons to come and collect those data and link the data that is collected through the forms with the implementation data so that additional research can be done. But locally, if we have to do that in the collection of implementation data, we need additional money, additional resources.

In conclusion, we had an agenda last year and tabled it. We wanted to do a longitudinal study of bilingual program students to present those data to the state legislature. We tabled it because we did not have staff, but we have plans to go back and do that within the next two or three years if everything comes out right.

Jesus Salazar
Los Angeles Unified School District, California

I am going to talk about the Russian revolution, an on-going statistical revolution, and Title VII programs. You'll see how these three themes are related as I make some recommendations for Title VII programs. I'm currently evaluating the Eastman Project for the Los Angeles Unified School District (LAUSD). It is a seven-year longitudinal evaluation study that follows limited-English proficient students from kindergarten through the sixth grade. Let me give you some background on LAUSD. It has roughly a quarter of million limited-English proficient students. LAUSD's LEP population alone would make it the fifth largest school district behind only New York City, Los Angeles, Chicago, and Houston.

I am not going to explain the Eastman Project curriculum, that's another story. Suffice it to say that the Eastman Project served as the basis of LAUSD's Bilingual Master Plan that was implemented districtwide in 1989. Prior to the Bilingual Master Plan, primary-language instruction was provided by para-professionals in more than half of the bilingual classrooms. Schools implementing Master Plan models similar to the Eastman Project could reduce the number of bilingual classrooms by as much as 33 percent.

The seven-year longitudinal evaluation study I am conducting has resulted in a Title VII Exemplary Academic Excellence Award. At the time I began this evaluation study I did not know that I was going to be conducting Title VII Grant research. I've been learning as I go. The 1991 Russian Revolution: Paradigm for Statistics.

There is a major revolution going on in Russia as I speak. Who would have imagined the radical changes now occurring after more than 70 years of communist rule. I use the second Russian Revolution as an analogy because a similar revolution is occurring in statistics. The revolution in statistics began in the 1960s and is occurring at a slower pace than the Russian Revolution, but it is a revolution nevertheless. Back in the 1920s a major political-paradigmatic debate took place among statisticians in the social sciences and applied statistics. The "party" that finally won decided to report statistical findings in terms of levels of significance.

For those of you conducting evaluation studies and applied research in educational settings, you know that most parents, teachers, and school administrators couldn't care less about the probability level of a study. Quite frankly, I find research that reports only levels of statistical significance very boring. Ultimately, this type of research does not tell you very much. Parents and the educational community want to know only two things: first, ***does the program work?***, and second, ***How effective is the program?*** Unfortunately, the "party" that lost the statistical wars in the 1920s was in favor of reporting statistical results in terms of program effectiveness. That is, rather than reporting that a study was statistically significant at the .05 level with 34 degrees of freedom, these statisticians were more concerned in showing that variable X was 86 percent more effective than variable Y in improving reading scores. Any parent or teacher will relate to a study that shows a program can help a child learn to read 86 percent better.

Title VII Grants and Measures of Program Effectiveness

The California Department of Education has two requirements for Title VII Grant research and evaluation applications. First, the traditional level of statistical significance needs to be reported. Second, program effectiveness relative to a comparison program has to also be reported. The State Department of Education essentially has you do a "Pepsi Challenge Test." I was very happy when I heard Ms. Sevilla discuss her major concern about integrating the ivory tower research community with the public community. I believe that reporting data in terms of the effectiveness of a program begins to address her concern. That is, the research community can continue to conduct multi-variate analyses with sophisticated research designs, yet the findings can be reported in terms of program effectiveness. We can all benefit from this type of research paradigm.

I conducted a three-year longitudinal study and performed multiple analysis of variance (MANOVA) statistics, yet I have been able to make this data meaningful to parents, teachers, and administrators by reporting the academic effectiveness of the Eastman Project. This approach to data analysis is so applied oriented that I have been able to present the positive longitudinal effects of bilingual instruction to parents in Spanish.

Two questions are always asked of me whenever I do presentations before parents and teachers. First, parents want to know, ***"Is my child learning English?"*** That's what they basically want to know. Second, teachers ask me, "Is the program I'm teaching in working? How effective is the program and why should I continue using this instructional model instead of the model that I was using?" These questions were common both in the initial phase of the Eastman Project implementation and in the initial period of the Bilingual Master Plan implementation.

EFFECTIVENESS OF EASTMAN PROJECT COMPARED WITH COMPARISON SCHOOLS AND DISTRICT

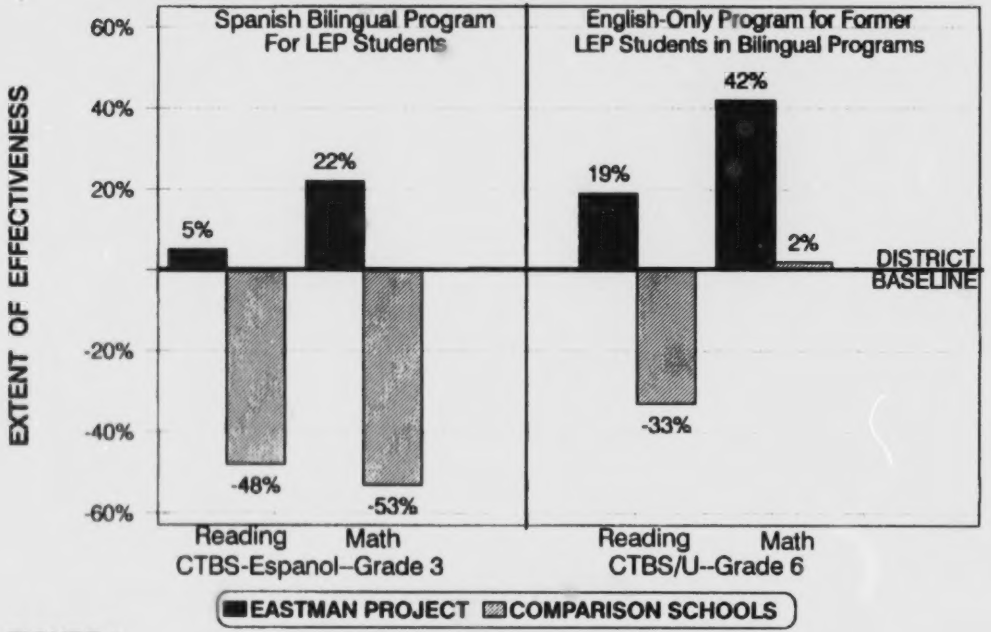


FIGURE 1

Third-Year Implementation (1988-89) CTBS Scores

Figure 1

Eastman Project:
An Effective Academic Excellence Program

After three years of implementation, third grade limited English proficient students who received three years of Spanish-language instruction in the Eastman Project were learning to read in Spanish 5 percent better than LEP students districtwide and 53 percent better than a group of LEP students who received three years of instruction in a comparison bilingual program (see Figure 1). LEP students at the Eastman Project schools were also learning Spanish-language math 22 percent better than the district baseline and 75 percent better than the LEP comparison group. In short, the Eastman Project is a more ***effective program*** for teaching LEP students in the Los Angeles Unified School District.

Though these findings of Spanish-language instruction are very encouraging, the key question for educators and parents is, ***how well does this knowledge acquired in the Spanish classroom transfer into English-only instruction?*** We have a preliminary answer, but a very encouraging one. Figure 1 also shows that the English Project has been successful in providing English-only instruction to students who previously received Spanish-language instruction. Briefly, the Eastman Project was 19 percent more effective in teaching English reading to former LEP students than the District English reading baseline. The project is also 42 percent more effective in teaching former LEP students to read in English than the comparison school English-only program. The Eastman Project was also 42 percent more effective than the District in teaching math in English-only classrooms, and 44 percent more effective than the comparison school program.

As I mentioned earlier, what makes these results exciting is that not only were these ***statistically significant findings*** but that the results were also presented in terms of program effectiveness. That is, every parent I've met would prefer that her/his child be enrolled in a program that is 19 percent more effective in teaching the child to read English. One of the recommendations that I have for Title VII grant applications and reports is that program effectiveness be presented graphically as I have in this presentation. This is the type of graph that you see in the Wall Street Journal. This data, as I indicated earlier, is based on a MANOVA analysis. Yet, when presented graphically and in terms of program effectiveness, the data become even more powerful. Again, let me emphasize the practical applications of this model. The findings of this model can be used in parent-teacher conferences to provide parents with information regarding the most effective programs for ultimately teaching English to LEP children. In the Los Angeles Unified School District, parents can choose whether or not to enroll their children in bilingual programs.

This information about program effectiveness can be provided to parents to facilitate their decision.

Recommendations for Title VII Evaluations

I want to identify four models for evaluating Title VII programs. These models are based on my search of the research and evaluation literature over the past two years. These models are listed in order of importance for reporting the effectiveness of instructional programs. That is, evaluation models should first and foremost highlight the effectiveness of an instructional program. If statistical significance needs to be sacrificed for the sake of evaluating program effectiveness, then so be it. The four models are listed below in their order of importance for Title VII evaluations:

1. A program is both educationally and statistically significant
2. A program is educationally significant and statistically non-significant
3. A program is statistically significant and educationally non-significant
4. A program is both educationally and statistically non-significant

A program is considered to be educationally significant if it is demonstrated to be a more effective instructional program when compared to another program. Studies exist where highly statistically significant findings were obtained but yet were basically educationally non-significant (e.g., one program was 3 percent more effective than another program). There have also been instances where an evaluation study did not reach statistical significance, yet one program was demonstrated to be 15 percent more effective than another program in teaching students to read English. Under the current research and evaluation paradigm, the case where statistical significance and educational non-significance is obtained is considered more noteworthy than the case where statistical non-significance and educational significance is demonstrated. However, for practical applications the latter case carries greater educational importance. As I mentioned earlier, most parents would rather have their child to be in a program that is 15 percent more effective in teaching English reading than a program that is only 3 percent more effective. Statistical significance be damned!!! After all, statistical significance is not able to teach Johnny or Juanito to become a better English reader!

Epilogue

Statisticians and researchers in the areas of meta-analysis and power analysis are leading the change in re-emphasizing the effect size of programs. Jacob Cohen, who is my statistical hero and the preeminent statistician today, was the first to emphasize effect sizes in the 1960s. Thirty years later statisticians in the social sciences and educational research are still reporting the majority of their findings in terms of statistical significance. However, with others such as the prominent Harvard psychologist Robert Rosenthal leading the way, measures of program effectiveness have become more common in the 1980s and 1990s. The revolution to report research findings in terms of program effectiveness is thus gathering momentum.

I want to close with one of my favorite quotes attributed to Henry Ford. Ford is quoted as saying that...***"if you can't write down your idea on the back of my business card, then you don't have a good idea."*** This reflects my sentiments regarding program evaluation and research. That is, many researchers do not present their evaluation findings concisely and to the point. Let me put it this way, which idea would Henry Ford have been most likely to be impressed with, to write on the back of his business card that the Eastman Project is 42 percent more effective in teaching English reading to former LEP students; or that the difference between the Eastman Project and comparison school reading programs is statistically significant at the .05 level as analyzed with a MANOVA with 75 degrees of freedom using a repeated measures design...?

Tomi Deutshc Berney
New York City Public Schools

Title VII evaluators are intermediaries. They stand between the providers of information and its consumers.

First, the providers of information: Project staff know they must report a great deal of information. While we understand that project directors frequently have an inordinate number of tasks to accomplish, there is information that only they can provide, and providing accurate information is a critical responsibility.

Understandably, projects are loath to transcribe data which they've already supplied to someone else; they resent searching for numbers that, it later turns out, no one needs to know; they are opposed to giving students unnecessary tests. Neither project personnel nor evaluators wish to place unnecessary burdens on students and teachers. No one should be asked to perform superfluous tasks. It is the obligation of the evaluator to make the evaluation process as efficient as possible. We, in New York City, are currently seeking to do this.

Second, the consumers of information: Evaluation reports are not found on paperback book racks; they are not read for pleasure. Consumers want to inspect the instructional and non-instructional data provided on past and current program participants. Consumers want to study the information about program activities and materials. They want to know what the impact of the program has been on student achievement in English or the native language where appropriate, and content area subjects. They want to discover whether the project has met its specific program objectives. The consumers of information justifiably expect the evaluator to assist them in accomplishing these tasks as quickly and efficiently as possible; they value clarity and conciseness.

The process and the product of Title VII evaluations are equally important. In both, we're learning as we're doing. We have not yet finalized either how we are going about collecting information or in what way we are going to report that information.

The process of evaluation should be as uncomplicated and efficient as possible; the product of evaluation should be clear and concise. In enuring that both the process and product are as they should be, the evaluator must perform a myriad of tasks: prepare forms which are easy to complete accurately and fully, collect the data and analyze it, integrate information from a wide variety of sources, and write and edit the report. Evaluators must go through these steps each project year.

Despite this, implementing a process that is uncomplicated and efficient and providing a product that is clear and concise are possible. The Office of Research, Evaluation, and Assessment (OREA) of the New York City Public Schools has developed a plan to implement the process of efficient data collection and has a prototype for a succinct, information-filled report. But neither the process nor the product is static; an effective evaluation system should be adaptable.

The process of gathering information is based upon two principles: First, avoid asking for information unless you know exactly how you will use it. And second, inside almost every open-ended question is a concealed closed-ended question. When possible, ask the closed-ended question.

Technology is the means which offers the greatest opportunity to streamline the process of collecting data. Because electronic records can be accessed and updated so much more easily than can paper records, electronic record-keeping must be the medium of choice. In New York City, we have taken steps to ease the burden of reporting data by utilizing an electronic system wherever possible.

A central computer maintains all citywide test scores; many offices can access these scores. Title VII programs, therefore, need not report their students' scores on the Language Assessment Battery, the instrument with which we measure proficiency in both English and Spanish, or on any citywide tests. Once we have a participating student's name and student identification number, we can automatically get pre- and post-test scores.

We ask project staff to complete a Student Data Form for each participating and formerly participating (now mainstreamed) student. Most of the background information required by Title VII comes to us on this form. When we receive the Student Data Forms, we enter the information they contain into the system. The following year, we preprint the Student Data Forms for continuing and mainstreamed students, using this background information already in the system. Project staff need only report any new information, including data on attendance and academic performance.

We used to ask that separate forms be completed in the fall and in the spring for high school programs. This year we developed a single form with fields for data for both semesters. Ultimately we would like to reduce the burden on project staff even further by retrieving individual attendance rates and course grades from the central computer files.

To learn about staff qualifications and program activities and materials, we developed a Project Director's Questionnaire (P.D.Q.). This is followed by a structured interview. We have tried to make both the P.D.Q. and the interview as efficient as possible. We ask for information that Title VII requires. We also ask for explanatory information to help the consumer of the information better visualize and understand the project.

The product, the evaluation report, is obviously at least as important as the process of evaluation. The report should be clear, concise, and filled with information. Title VII regulations may dictate the content of the report, but the evaluator can choose the form. In New York City we have developed what we call a profile format. In developing the format, we first worked from a list of Title VII regulations and made sure that each regulation was covered in the report. We then held focus groups and spoke to project directors and personnel from the State Education Department of New York. They recommended changes that would make the reports more useful to them. These recommendations were extremely helpful.

The profile format has gone through a number of changes. It's still going through changes. We are learning as we're doing. For an evaluation system to be efficient and effective, it must be adaptable. It can't be static because your populations change and your priorities change. It is very important that you change the evaluation report to meet the changing needs.

The report has two parts: the Extract and the Program Assessment. The Extract is not a summary. It presents the salient points, the most important points, the items which we were asked to state on the first page. Since there is too much information to state on one page, this section of the report is contained in the first two pages. The Extract contains information on the funding cycle, sites, enrollment, background of the students served, admission criteria, programming features, and strengths and limitations. It also gives OREA's conclusions, including which objectives were met and which were not, and recommendations.

The funding cycle indicates what year of funding the project has just completed. The sites section lists the sites in the project, the grade levels included, the number of students participating in the program at each of the sites. Student background lists the number of students by native language and country of origin. We include here information on how many years of education the students had, on the average, in their native countries, how many they had in the United States, and what proportion of students were eligible to participate in the federally-funded free lunch program. Admissions criteria includes any criteria the program uses for program participation.

Programming features, strengths, and limitations of the project presents those and states what objects the project met, which it did not meet, and for which it provided no information. Reasons why a project may not have achieved a certain objective are given later. In the recommendations section we frequently recommend exploring reasons why objectives were not met, suggest that objectives be modified to make them more realistic, or suggest ways of meeting objectives or providing data. We try very hard not to recommend things that necessitate the expenditure of additional funds.

The Program Assessment is the major part of the report. Its sections are: staffing, implementation and outcomes, services to students with special academic needs, mainstreaming information, and a case history.

Staffing lists the title, highest educational degree, and language competencies of the Title VII-funded staff. Other staff who work with project students (teachers, for example) are described in aggregate.

The second section, implementation and outcomes, is structured around the objectives of the project. For each objective, we report relevant activities, the evaluation indicator used, and a summary statement as to whether the project either met or did not meet the objective. The section may include teaching techniques and materials, or the latter may be listed in an appendix. Whether or not there are objectives concerning attendance and dropout rates, information on those is presented in this section.

The part of the report where objectives are presented and discussed is critical. Objectives define the direction of a project. A well-stated objective helps the project tell the world how good it is. A well-written objective clearly states who is expected to accomplish it, what the expected performance is, and when the accomplishment of the objective will occur. Unless objectives fulfill these criteria, they should be considered unacceptable.

The third section shows statistics on students with special academic needs. This includes data on students referred to special education, to remedial programs, to programs for the gifted and talented, and how many students were retained in grade. We include here the linguistic competencies of the school staff who evaluate students for these programs. We look both at the number and the percentage and we attempt to compare the current year's data with those of the previous year to see if there has been any change. Has the project made a difference?

The fourth section gathers information on students mainstreamed and the number of graduating students planning to enroll in postsecondary education institutions. This section concludes with a report on the academic progress of former project participants who have been mainstreamed.

The case history, though not required by federal regulations, is the fifth section. It suggests the program's impact on individual students. The case history gives the consumers of the information the ability to visualize and to understand the project that we are evaluating. In the case history, we carefully maintain confidentiality.

We feel that the Profile Format will meet the needs of those involved in Title VII projects as well as state and federal officials who review the achievements of these funded programs. With further experience and continued feedback, we will continue to refine the form. We have added a page to the profile format that explains how we gather the various kinds of data as well as the statistical procedures we have chosen to use.

One additional advantage of a clear and focused format is that it facilitates the preparation of academic excellence applications. The profile format provides information in a way which simplifies the task. It can be easily determined whether a project may be considered to be exemplary in any area. A description of participating students and staff, program activities, academic and non-academic achievements, and the degree to which the project met its objectives are presented clearly and concisely.

Evaluation should be a high priority. We feel that this priority should be reflected in the number of points allotted to evaluation on the Title VII grant application, and more specifically, that it be required that objectives be clearly stated, measurable, and realistic. An evaluator can easily assist a proposal writer or a prospective project director develop well-written objectives. The New York City Public Schools Office of Research, Evaluation, and Assessment routinely offered this service. Those who write proposals and those who approve them should place greater emphasis on objectives and project evaluation. It is hoped that the reauthorization of Title VII for 1993 will address this issue.

Three things are necessary to improve the quality and value of Title VII evaluations. First, the process of evaluation should be as uncomplicated and as efficient as possible; this is becoming more attainable with electronic record keeping. Second, the product should be informative, clear, and concise; these are the goals of the Profile Format. Finally, objectives should be well-formulated, clearly stated, measurable, and realistic.

In order to maximize program effectiveness, it is of the utmost importance to prioritize evaluation. In doing so, it is essential that we continually assess both the process and the product of evaluation and modify them as necessary.

Discussion of Panelists Balu, Salazar, and Berney's Presentations

Robert Martinez
University of New Mexico

What's being passed out are the Title VII regulations with which all LEAs receiving Part A funds are required to comply. After they are distributed, I would like to talk about them with you and the group. This is basically what Alan Ginsburg had called "the laundry list" on his first day presentation. This has been our laundry list, so-called, for the last six years. The reauthorization of Title VII will be coming up very shortly.

It is now time to do, in testing terms, an item analysis with these Title VII evaluation regulations. We need to look at those regulation items -- those we need to keep, those we need to revise and, most definitely, those we need to exclude.

What I'd like to do at this point is to address six specific regulations since we don't have time to address them all, but six that I feel have credence. They had a lot of problems in being addressed or not being addressed in the field. I would like to address them by first stating the regulation and then having my colleagues address how they work with that regulation and their respective school district.

First one. Under 500.50(b)(1), a grantee's evaluation design must include a measure of the educational progress of project participants when measured against an appropriate non-project comparison group.

Before my colleagues respond to this, I would like to say that we've looked at these evaluation requirements in addition to regulation requirements required by Chapter 1, Special Education, Indian Education, Migrant Education, and we find that these are the most comprehensive and the most stringent. However, they do provide for program improvement which our programs are all about. With that, I'd like to work eastward starting from the west.

Jesus, would you mind addressing the requirement that a grantee's evaluation design must include a measure of the educational progress of project participants when measured against an appropriate non-project comparison group? Could we just limit our discussion to maybe a minute for individuals?

Jesus Salazar

You caught me off guard, but I can try and answer it. I think I was mentioning in my presentation that we actually have two kinds of comparison groups with district baselines, but we also have a group of schools very similar in demographics to the project schools. We just use traditional California Test of Basic Skills test scores, California Assessment Test Scores, but the extra little twist that I add, and that's the one I mentioned, is effect sizes. I also report the effect sizes for the comparison schools that must report.

Robert Martinez

Is there a certain design that you use most often?

Jesus Salazar

I have conversion tables for anyone who is interested. Any statistics test, any research design you use, after you've run the statistical analyses, you can convert that analyses in effect by a multiple analysis of variance (MANOVA), one way, two way, two by four, you name it, and I can provide an effect size.

Raj Balu

We have in Chicago different kinds of comparison groups identified for Title VII programs. One is within a school, and one set of comparison groups within school groups as well as between school groups. We have Title VII programs and, within the same school, we have transitional bilingual education programs mandated by the state for students who are not receiving Title VII services. So those two groups can be compared, and we also have Chapter 1 students and English monolinguals, English students who receive English instruction only, not even transitional bilingual education services from the state. So there are four different groups that are used in our data analyses: Title VII group; the state bilingual program without Title VII; students with no bilingual program, that is English only instruction; and, finally, Chapter 1 students who are bilingual and receiving services under Chapter 1 and bilingual services. They are receiving two sets of services very similar to Title VII.

Let me qualify one more item. The students who are in Title VII programs receive two kinds of services -- the state bilingual services and the supplementary services from the Title VII programs. Under the developmental program, we are thinking of using the grade-design and follow them for up to eight years; that is, we follow different students who get into the program at different times.

Tomi Berney

In New York, we really have a different set of circumstances. In New York City, every student who is limited English proficient is entitled to receive supplementary services, so we really can't do a control group design, an experimental group/control group design. What we use is a gap reduction design and, instead of using an equivalent control group, we use the norming sample, the group on which the language assessment battery was normed.

Those of you who have our sample report, on the bottom of page 3, we've specifically discussed this, how we use the gap reduction and the way in which we use it is we talk about normal curve equivalence (NCE). When you're talking about NCEs, you are assuming that there will be no gain, there will be zero change in NCE from year to year; we use Spring-to-Spring evaluation. When there is a gain, it means that the participating students are doing better than we would have expected the group on which the test was normed to have done. Assumedly, they will do better than just one NCE better, we would hope for at least five but, in any case, it is a gap reduction design that we are using.

Robert Martinez

The second regulation I will address is 500.50(b)(2)(ii), reliability and validity of the evaluation instruments and procedures. The evaluation instruments used must consistently and accurately measure progress toward accomplishing the objectives of the project, and they must be appropriate considering factors such as the age, grade, language, degree of language fluency, and background of the person served by the project. I'm particularly interested in addressing those populations where standardized tests are not available. What do you use at that point?

Raj Balu

First, let me answer by explaining the kind of standardized test that we are using now in the Chicago public schools, the first one initially that we used for students admitted to the school programs. It is a functional language assessment instrument. It's a simple instrument that is testing the auto language skills of the students, and we have questions about the validity of that instrument, and we are in the process of revising that.

The second group of instruments is used during the initial enrollment of the students into the bilingual program; this is their language assessment scale, which is a standardized test; it is being used now. One of the concerns we have is that it is time-consuming, mainly the auto component of that particular test is time-consuming.

The reading and writing we are going to continue to use for some time.

During the spring, we are using the Iowa Test of Basic Skills for the first grade through the eighth grade level and then the Test of Academic Progress in the high school grades. These are the same tests that are used for citywide testing, so that we will have comparable data for viable programs and students who are in different kinds of program situations.

When tests are not available, an example of a situation I can say right now is that the auto language component assessment has become a little difficult for us. The language assessment scale takes about 45 minutes for each child. We have 45,000 students, and we are required to test every one to assist the progress of the students. Currently what is being done is that the teacher is asked, read to the children; develops oral language proficiency from one through five. That's all that's being done, and we are trying to bring into the system the ways that they rate the children and ways of increasing the validity and objectivity of this particular process.

Tomi Berney

The language assessment battery that we use in New York is reliable, and it's valid. In fact, it was just renormed this current year. It has four components -- listening, speaking, reading, and writing. When students come into the system and are initially tested, let's say in the fall or in the middle of the year, they take what's called a short version of it. This does not include reading, and it doesn't include speaking. But on this score we determine whether a student is limited English proficient or not.

Then when the spring testing takes place, we give the full lab which again really does not include the speaking sub-test, but that's something else. It includes the listening, the reading, and the writing components, and we can measure from year to year how the student is doing. We do use the speaking component for something else; it's not that we ignore it; we just look at that score separately.

The norms that we use, the lab itself, are the English proficient norms. We're not using limited English proficient norms. A student, to show growth, really does have to gain in skills. However, it is reliable, it is valid. In some cases, we do use other citywide tests depending on whether the student has been in an English speaking school system for at least two years. Where we run into problems is testing students in languages other than English. In Spanish, we have no problem because the lab is in Spanish also. We are in the process right now of developing a Chinese reading test, which we really don't have at this point. The other major language group in New

York is Haitian Creole, and we don't have a test there. What they have been using in those cases are either teacher developed or district developed tests.

Jesus Salazar

There is a section in Los Angeles, actually it's the Hollywood section, which is better known for the stars' walk of fame, etcetera, that has 60 languages represented. So you can imagine the situation teachers in schools are facing. We rely to a large extent on the Student Oral Language Observation Measure (SOLOM), which is teacher assessment. It takes about 20 minutes, and teachers usually give it about a month after they have had a student in their classrooms. That's the measure we use to identify a student's level of English proficiency or lack thereof.

We are also in the process of developing an Asian language curriculum for the district. We have, I believe, 40,000 limited English proficient students. We have a lot of Armenians, a lot of Russians, and overall the district has about 88 languages represented. We're in the process of trying to address as many as we can. We rely on the CPPS as a measure for transitioning students from native language instruction into English language instruction. They have to meet the 36th percentile, so we rely a lot on norm testing once again, English. Prior to that, as I mentioned, we rely on teacher observation measures to identify for placement in programs.

Robert Martinez

The first day of the conference was focused on alternative assessments for performance-based assessment. Is it now time for that to be included in Title VII regulation requirements? I'm not going to turn it over to you for an answer at this point, I'd like to continue the other ones. But it is food for thought in that the authorization will be up. You may want to consider that and address it with the appropriate personnel.

Under 500.50(b)(3)(i)(B), evaluations must provide information on the academic achievement of children who were formerly served in the project as limited English proficient, have exited from the program, and are now in English language classrooms. How has Chicago addressed this requirement?

Raj Balu

The current exit criteria for LEP students is that they need to be at the 50 percent cut off point. That is, they should perform at the fifth stanine before they are ready to exit from the program. There is a conditional exit that is the fourth stanine and that is a recommendation. All of these are on the Iowa Test of Basic Skills or the Test of Academic Progress, depending on the grade level of the student. Now, once the student exits, we follow the student for a year and, if the teacher recommends this child be brought back into the program, then we bring the child back into the program. THIS IS REQUIRED BY THE STATE LAW.

Robert Martinez

What about the children who have exited from this program?

Raj Balu

The exited students, especially the conditionally exited students, are followed for at least one year and, if necessary, brought back for support services, not bilingual education but transitional programs.

Tomi Berney

In New York, I mentioned before that we distribute student data forms. One of the student data forms is for previous, now-mainstreamed program participants. On this form, the school must give information concerning class grades, sometimes test scores, such as New York State Regency Examinations, or Regents Competency Test, and attendance data. In this way, we are able to follow the student. The one problem we have is that the schools do not like to bother filling out information; these are no longer program students, and it's very hard to find people to give us this information. That's what I was saying before, to ensure that we get the information and get it into an electronic system would be much to our benefit. I'm sure we miss a lot of students. We don't get all the mainstream students.

The other problem is when a student goes from school level to school level -- from elementary to middle school to high school -- we lose track of that student. Ultimately, New York City is supposed to be on a computer system called Automate The Schools (ATS). It's not in all districts now so we can't access this information everywhere. As far as I know there are not yet any plans to do this in the high schools. Once every school is on computer, it will be very simple to get this information on any student with the name and ID number of the student. We can't get it now. We have to rely on school personnel to give us these data. We tried to make the forms as simple as possible, but we can't force them to do it. That's the problem involved with it.

Jesus Salazar

In the Los Angeles Unified School District, all of the elementary school students are on a computerized database. We have gotten to the point where even if they move to another school within the district we can follow them because of ID codes. We have half of the junior high school students on the computerized database and half of the high school students on the same system. By the end of this academic year, 91-92, we should have all students entered into the computer database. What we're doing is following former LEP students who have exited from a bilingual program, not just Eastman but any of the other eight bilingual programs we have; we can follow their academic progress until they leave the district.

I'm currently doing a follow-up study of the original Eastman elementary school -- that's what the Eastman project is based on. It started back in the 81-82 school year. I'm following the different cohort group of students. Most of them are 10th, 11th, and 12th graders currently. I'm doing a 10-year follow-up to see if the academic achievement has been sustained over a 10-year period.

Robert Martinez

Two more, if you would, and then I'll open up to questions. Under 500.51(f) Title VII grantees must collect information on the specific activities undertaken to improve pre-referral evaluation procedures and instructional programs for LEP children who may be handicapped or gifted and talented. Chicago, how do you deal with that?

Raj Balu

We have two different departments that handle the education program of these children. One is the special education department. It is the primary one in terms of identifying and following through the referrals and assessment of the children as needed. The other is bilingual education, in which the language and cultural education department gets involved. There are two tiers of the gifted program. One is across the board for all students, and LEP students are also eligible to participate in that particular program. In addition, last year, we devised a Spanish-English gifted children's program, and that begins this year. Once this becomes a practical program in terms of planning, implementation, materials-development and so on, then the idea is to expand this to other languages as well as to additional schools in the Spanish language.

Jesus Salazar

The Los Angeles Unified School District also has a referral, teacher-test situation. Unfortunately, some of the coordinators at the school sites were of the opinion that students cannot be classified as gifted until they learn English. That's unfortunate because we've seen cases -- we've caught cases like that. We now have a system where we can bypass coordinators who feel that way. We have an office of special education that identifies gifted students in all languages, Spanish, Korean, and the other 86 languages that we have in the district. Everything is maintained on a computerized database so that at any one point we can keep track of the gifted students.

Tomi Berney

Fortunately, in New York all programs are open to LEP students -- gifted, talented, or remedial programs or any others. It is specifically stated that LEP students are eligible. As far as that certification for special education, every district has a school base support team and a committee on special education. Here there is a school psychologist, an educational evaluator, and a social worker who make a recommendation as to whether a student belongs in special education. It is hoped that, in cases where the student is limited English proficient and where the parents really don't communicate in English, the person who either tests the student or speaks with the parents is able to communicate in the language of the student and/or the parents.

In addition to the fact that all gifted and talented programs are open to LEP students, we have specific programs for the gifted and talented LEP students. I evaluate at least 60 programs so we have some of everything in New York City.

Robert Martinez

Last one. 500.52(c)(5) asks grantees to report on the extent of educational progress achieved through the project measured, as appropriate, by changes in the rate of student enrollment in post-secondary education institutions. How does Chicago handle that?

Raj Balu

About two or three years back, there was a problem of LEP students who graduated from high school taking the ESL One and ESL Two courses and the basically many, many needed English courses to get admitted into the colleges and the universities. They needed to take additional courses. In the last two years we had a committee task force working with the university and we have resolved that particular problem. There are different standards set now under the ESL for these students who graduated from high school to get into the program and we follow some of the students -- not in detail.

Tomi Berney

The question really concerns how do we follow this. What we do, we ask a question on the student data form. First we find out what grade the student is in, and we assume 12th graders are graduating from high school as long as they're not being retained in grade and they have not dropped out, of course. We then ask a specific question, does this student plan to enroll in postsecondary educational institutions and we just compare the two pieces of data and we find out what percentage of students who are graduating from high school are going to be entering postsecondary institutions.

Jesus Salazar

In the Los Angeles Unified School District, I'm with the bilingual unit of program evaluation and assessment. We have five different units. The research unit follows students who go into higher education, and they keep track of as many students as possible. What I feel we are going to be doing in the near future is to follow through on students in different types of bilingual programs within a district who have exited into English only programs and have gone on and graduated and gone into college. We're probably doing that type of longitudinal study, and I mentioned earlier that I'm doing a study, a 10-year follow-up of the original Eastman elementary school students. Some of them, the cohort group that was in the fourth grade, are currently freshmen, those that have continued in college. I believe that I'm going to be asked to do another study to follow up that group of kids, so they're probably going to have to work for me when I get back.

Evaluating the Mathematics Education of Limited English Proficient Students in a Time of Educational Change

Walter Secada
University of Wisconsin, Madison

Bilingual-Education-Program Evaluation: Current Practice

Program evaluation and related research have come a very long way from the quasi-experiment as formalized by Campbell and Stanley (1966; Cook & Campbell, 1979) to where program evaluation is now seen as having many functions, as being grounded in a range of theoretical positions, and as drawing from a variety of possible methodologies (Cook & Shadish, 1986; Cronbach, 1980; Lindblom & Cohen, 1979). In practice, however, the evaluation of bilingual education programs has not strayed very far from its original, basic question: Does the program work better than not having the program? Or, Does the program work better than having a particular, alternative program?[1] At one time, the law that provided federal funds for bilingual education required districts to compare performance by students who were in the program to performance by students who were not. This has been the minimal question that the evaluation of federally funded programs should try to answer.

Regardless of this narrow focus in bilingual-education-program evaluation, it has been de rigueur to bemoan the quality of evaluations that have been produced by federally funded projects. On this point, sympathizers, critics, and people who are neutral about bilingual education all seem to agree (Baker & De Kanter, 1983; Boruch & Cordray, 1980; Willig, 1985).

Elsewhere, I have speculated on some of the reasons for these two problems with current practice in bilingual-education-program evaluation: (a) the failure to move beyond a very narrow set of questions to other questions that are no less interesting and that are, in many ways, more important to local stakeholders; and (b) the failure to meet technical standards of rigor. This is not to claim that there have been no advances in the field. New models for program evaluation, the best known being the gap-reduction model (Tallmadge, Lam, & Gamel, 1987a, 1987b) have been developed. And, federally funded large-scale evaluations of bilingual education have come a very long way from the AIR Report (Danoff, 1978; Danoff, Coles, McLaughlin, & Reynolds, 1977-1978) when there were no efforts to ensure prior-to-treatment comparability of the comparison groups or to document the fidelity of programs to their descriptions. Though many people on all sides of the "effectiveness debate" may not be satisfied with the conclusions of the Longitudinal Study (Ramirez, Pasta, Yuen, Billings, & Ramey, 1991; Ramirez, Yuen, & Ramey, 1991; Ramirez, Yuen, Ramey & Pasta, 1991), it did document fidelity of treatment and it ensured comparability of groups. That study has also moved the field of program evaluation forward in many other ways: for example, it served as a testing ground for new statistical methods like hierarchical linear modeling (HLM).

Regardless of these developments, there have been at least two constant foci of debate in bilingual-education-program evaluation on federal and local scales. These debates have been around the goals of the program and the sorts of evidence that evaluation can provide.

Bilingual Education Program Goals

Over the years, there has been quite a bit of debate about the range of goals that are appropriate for programs of bilingual education. At the start of the federal funding initiatives, from the late 1960s and into the early 1970s, this debate was couched in terms of two poles that used a variety of terms: assimilation and monoculturalism versus pluralism and biculturalism, the development of English and of English literacy versus native language maintenance, the development of balanced bilingualism, and biliteracy (Andersson & Boyer, 1978; Mackey & Beebe, 1977; Stein, 1986). Pluralist views on the purposes of the program came under concerted attack almost as soon as they were articulated (e.g., Epstein, 1977), and the AIR Report (Danoff et al., 1977-1978) found that such programs did not enhance elementary-school Hispanic children's achievement (in English) better than if there had been no program in place.[2] The mid 1970s was a time of retreat from the purported excesses of the late 1960s, among them cultural pluralism. Thus, the federal-funding program has come to be sharply defined around two goals: the development of English language skills by LEP students and the development of their academic skills so as not to fall progressively behind their English-proficient peers (Secada, 1990a; Stein, 1986).

In recent debates about the goals for bilingual education, some authors have written as if the federal government were mandating a single approach or as if the only goal of the program were to develop English-language skills (see Baker & de Kanter, 1983; Government Accounting Office, 1987). Eleanor Chelimsky, director of the Program Evaluation and Methodology Division of the GAO, argued against this overly narrow specification of goals when she testified before Congress:

> To say, first of all, that there is a method mandated in the act when, in fact, the act says we will use native language to the degree that is necessary -- that is all it says....The same is true for the business of the two goals. To say the act only has one goal, teaching English, is to ignore the other goal of the act which has to do with keeping people up to date in all their subjects. (Reauthorization..., 1987, p. 30)

Other writers have argued that, granting the transitional and assimilationist bent of the above goals, they are too modest. For example, in reviewing the research literature on culturally-diverse populations and mathematics achievement, I noted that there should be at least three goals for intervention programs such as Chapter 1 and bilingual education: (a) improved achievement (beyond what would have occurred without the program); (b) a closing of the achievement gap between the population of interest and the so-called mainstream; and (c) long-term effects wherein the gap, once it is closed, remains closed (Secada, in press). The first of these goals is clearly a goal for bilingual education. The second appears, at least tacitly, in status studies that report the mathematics achievement of diverse learning populations compared to one another (reviewed in Secada, in press), in the gap-reduction model (Tallmadge et al., 1987a, 1987b), and in research designs like that of the Longitudinal Study (Ramirez, Pasta, Yuen, Billings, & Ramey, 1991). Though the third goal has been an explicit part of longitudinal studies like the Sustaining Effects Study and other evaluations of Chapter 1 (Kennedy, Jung, & Orland, 1986), I have been unable to find any evidence that the third goal -- long-lasting closure of the achievement gap -- has been considered in the design or the evaluation of bilingual education programs.

Orum (1983) argued for long-term and for nonacademic goals in bilingual education: reduced dropout rates and an increase in successful school-completion, transition from high school to postsecondary education or to the workplace, and staying at grade level. Christina Bratt-Paulston (1980) also has argued that the goals for bilingual education should be long range and that they should include out-of-school outcomes. Among her recommended indicators of success are:

> employment figures upon leaving school, figures on drug addiction and alcoholism, suicide rates, and personality disorders, that is, indicators which measure the social pathology which accompanies social injustice rather than attempts at efficient language teaching -- although programs are that too (p. 41).

There is wisdom in these recommendations, not only because of the vision of schooling that they propose but also because the payoffs for programs such as bilingual education may, in fact, be long term.

The case of Head Start is illustrative in this regard. Initially, Head Start's goals were short-term and cognitive. And on those grounds, that program fell into deep trouble, much as has been the case for bilingual education. It was on the basis of the long-term, and especially the out-of-school, outcomes of Head Start that it finally achieved the widespread social science and political support that it currently has (Stallings & Stipek, 1986; White & Buka, 1987).

Measurement of Goals

The measurement of bilingual-education-program goals, especially of its academic goals, usually has been translated to mean academic achievement. Typically, as in the case of Chapter 1 evaluations, reading and mathematics are the subjects for which academic achievement information has been gathered.

There have been some debates about the language of the achievement tests that are administered. Some writers have argued that, since the ultimate goal is for students to function in an all-English-speaking setting, achievement should be measured only via English language tests (Baker & de Kanter, 1983; Danoff et al., 1987-1978). Others have argued that, even though the eventual goal is to function in an all-English setting, achievement in either language should be measured in order to get as complete a picture as we can of students' actual learning of content (Willig, 1985; Ramirez, Pasta, et al., 1991). As a proxy for achievement, large-scale studies involving bilingual populations also have used indicators of engaged time on task (e.g., Tikunoff, 1985).

In mathematics achievement, the AIR Study (Danoff et al., 1977-1978) found that only in fourth-grade mathematics achievement did children enrolled in bilingual programs outperform children who were in neighboring school districts and were not enrolled in such programs. In their narrative review of the bilingual-education-program evaluation research, Baker and De Kanter (1983) found uneven effects of programs on mathematics achievement. However, in her meta-analysis of a subset of the Baker and de Kanter studies, Willig (1985) found that children enrolled in bilingual programs outperformed control children on standardized tests of mathematics achievement, whether those tests were administered in English or in Spanish. Interestingly, Willig also found that the better the technical quality of a study -- e.g., if it used random assignment of students -- the more likely it was that the evaluation would show favorable results for the program.

Ramirez and his colleagues did not conduct a direct comparison of various program models[3] against each other due to confounding school-level with program-level effects. In an effort to circumvent those problems, Ramirez et al. compared how well students in each program performed against the norming populations for the standardized tests that were administered. Between kindergarten and first grade, children in all three programs grew more quickly on an English-language standardized test of mathematics than did the norming populations (see Ramirez, Yuen, & Ramey, 1991, Figures 7, 8, & 9). Between first and third grades, children in all three programs kept pace with the norming population (Figures 10, 11, & 12).

Next, Ramirez and his associates compared the growth in mathematics achievement among students who were enrolled in late-exit bilingual education programs, and had experienced different amounts of their native language (Spanish) over the course of their elementary school years. Students who experienced the most substantial and the most consistent use of Spanish began below national norms but grew the most in mathematics achievement.

> Students in site E, who were provided with substantial instruction in their primary language and a slow phasing in of English instruction over time, consistently realized the greatest growth in mathematics skills, faster than [the] norming population. Students in site D, who were exposed to a consistent proportion of instruction in their primary language (approximately 40 percent), realized growth in mathematics that was equal to [the] norming population. Noteworthy is that after covariates were considered, there was no difference in achievement of students in sites D and E, although students in site E had more stress in their environment and fewer resources than site D students. (Ramirez, Yuen, & Ramey, 1991, p. 33, emphasis added)

In other words, even though the students in one site lived in greater poverty and experienced more of what Ramirez (in personal communication) has termed the stresses of urban life (e.g., crime), they exceeded the norming population's growth and kept pace with a relatively more advantaged population. The tenor of Ramirez et al.'s observations leave little doubt that they ascribe this to the students' receiving substantial amounts of instruction via their native language, Spanish. Consider their observations about the third school in this sample:

> It appears that students in site G who received about 40 percent of their instruction in their primary language in kindergarten and first grade, but were then abruptly moved into almost exclusive instruction in English (comparable to that provided to early-exit and immersion strategy programs), experienced a marked decrease in growth in mathematics skills over time relative to [the] norming population. It seems that these students lost ground...paralleling what is commonly observed for disadvantaged students in the general population. (p. 33, emphasis added)

Surprisingly, though Ramirez and his associates collected achievement data via the students' native language of Spanish, they failed to report aggregate achievement data in Spanish and they did not analyze those data as they analyzed their English-language achievement data.

Thus, the best evidence that we have at this moment suggests that the use of children's native language -- at least for Spanish-speaking children -- for instruction in mathematics is more efficacious than instruction all in English. Moreover, the Ramirez study suggests that the more substantial and consistent the use of a child's native language during the primary-school years, the greater that child's growth will be -- up to the point where the gap between LEP and an English-proficient norming population actually decreases.

Omitted from most bilingual-education-program evaluations are other indicators of academic growth and whether or not, on those indicators, LEP students function similarly to their English-proficient peers. Continued course taking in mathematics should be one such concern (Chipman & Thomas, 1987; Oakes, 1990a, 1990b). Though achievement is important, the continued taking of mathematics courses is at least equally important since, regardless of achievement, one cannot take advanced courses without having taken earlier courses. In their study of the determinants of mathematics course taking by various ethnolinguistic populations in the High School and Beyond (HSB) data base,[4] Myers and Milne (1982) found differential patterns of course offerings and of course taking by high school males and females. We need to understand the reasons for such patterns and what we can do to encourage LEP students to take more mathematics courses in high school.

One reason that most program evaluations fail to attend to non-achievement indicators may be that most bilingual education programs are in elementary school where everyone takes the same mathematics course -- arithmetic -- and course taking does not seem to be an issue. But by junior high school, course taking is becoming optional and, beginning at these grades, it should be (but has not been) a concern.

The Taken-for-Granted in Current Evaluation Practice

Compensatory education was established with the idea of providing students the experiences and the skills that purportedly had been denied to them because of their culturally or linguistically impoverished upbringing (Kantor, 1991; Stein, 1986). Consistent with this belief, evaluation did not question the nature or the quality of curriculum or instruction that these students received. Curriculum and instruction were assumed as given.

Over the years, many writers have rejected such notions of deprivation that undergird the Great Society's Compensatory Education thrust (Kantor, 1991). But the programs that grew from that thrust and many of the assumptions that undergird those programs (and their evaluations) persist.

The Mathematics Curriculum

This general acceptance of the school mathematics curriculum is reflected in current bilingual-education-program research and evaluation practice. I have never seen efforts to document whether or not curricular objectives or materials are different for LEP versus mainstream students. In my own informal observations, however, I have noticed that, when a program for LEP students assumes the responsibility for the mathematics instruction of LEP students, the curriculum is very much focused on computational skills, and instruction tends to be individualized seatwork on pages and pages of worksheets. Mathematics instruction for Chapter 1 students (Kennedy, Jung, & Orland, 1986) or for students enrolled in low track courses (Oakes, 1990a, 1990b) can be similarly characterized.

Efforts to adapt the mathematics curriculum that LEP students receive have come about, mainly through content-based, English-as-a-second-language (ESL) approaches. The goal of these efforts is to develop English language skills through student engagement in mathematics, science, and social studies (Cantoni-Harvey, 1987; Crandall, 1987; Mohan, 1986). These approaches include a structural-linguistic analysis of what has been termed the mathematics register, and they tie that analysis to recommended goals for combining the teaching of mathematics with the teaching of English (Crandall, Dale, Rhodes, & Spanos, 1987; Dale & Cuevas, 1987; Spanos, Rhodes, Dale & Crandall, 1988).

O'Malley and Chamot (1990) have conducted an extensive series of studies documenting the learning strategies used by second-language learners as they learned their second languages (English being among the languages of interest), and for in-school populations, as they learned academic subjects such as mathematics. The results of their studies have included curriculum materials (Chamot & O'Malley, 1988) that try to combine second-language-learning and mathematics-learning.

For both of these approaches, content-based-ESL and language-learning-strategies, mathematics remains constant. There are no questions about its goals and objectives, nor about the adequacy of extant curriculum to meet those goals.

Students' Mathematics Learning and Thinking

Both content-based-ESL and learning-strategies approaches for teaching LEP students might help provide insights into how bilingual students learn mathematics. They entail at least tacit critiques that current mathematics teaching fails to match how people learn a second language and that it may not match -- how LEP students actually learn mathematics. For example, one might use the structural linguistic analyses provided by Crandall and her colleagues (Crandall, Dale, Rhodes, & Spanos, in press; Dale & Cuevas, 1987; Spanos, Rhodes, Dale, & Crandall, 1988) to argue that the reason LEP students do not achieve as well in mathematics as their English-proficient peers is that they lack knowledge of the mathematics register (Orr, 1987, makes a similar claim for students who speak Black English Vernacular). Unfortunately, there is no evidence that English-proficient students have any better grasp of that same register. Were such evidence forthcoming, it would provide a linguistic basis for looking at the school mathematics curriculum.

Carpenter (1985) has argued that, as early as first grade, the school mathematics curriculum ignores the rich stores of informal mathematical (as opposed to linguistic) knowledge that children bring to school. That mismatch, according to Carpenter, lays the foundation for widespread failure and disenchantment with mathematics among older children. Unlike other claims about children, Carpenter's is an argument based on competence -- children enter school competent in mathematical reasoning; the schools ignore that competence; and hence, the typical result of schooling is incompetence in mathematics. A similar case might be built vis-à-vis bilingual students.

There is a common folklore that bilingual students cannot solve arithmetic word problems and that the best we can hope for is to provide them with key words and other tricks for solving such problems. But in my work, I have found that first grade Hispanic bilingual children can solve many of the same word problems that have been used in studies involving monolingual children (Secada, 1991a). Moreover, I have found that competence in solving arithmetic word problems varies as a function of children's proficiency in the language in which they are assessed and also in degree of bilingualism when that language proficiency is assessed qua mathematical language.

Finding Out/Descubrimiento (FO/D; De Avila, Cohen, & Intili, 1982; De Avila, Duncan, & Navarrete, 1987) seems to have been developed along lines that combine what was known about concept formation and second language learning. Like Carpenter's argument, it is based on the tacit assumption that LEP students have more capacity than they are usually credited with. But FO/D extends

Carpenter's argument to include both academic and linguistic competence. Cheche Konnen (Warren & Rosebery, 1990) is another recent effort to identify and to capitalize on how bilingual students learn both content (in this case, science) and language.

Instruction in Mathematics

When considering the quality of instruction, most bilingual-education-program evaluations have focused on the role of the native language or on the role of instruction in developing students' English language skills. In their review of research on the teaching of bilingual learners, Fillmore and Valadez (1986) considered whether mathematical knowledge would transfer from a child's native language into English and when mathematics -- the universal language -- could be taught all in English. The Longitudinal Study documented how teachers dominated classroom conversations and how they asked very low-level questions when they tried to bring their students into a conversation (Ramirez, 1986; Ramirez, Yuen, & Ramey, 1991; Ramirez, Yuen, Ramey, & Pasta, 1991). Ramirez et al.'s critiques of that instruction were based on how such settings are less than optimal for the development of English as a second language. They said nothing about how such settings are also deadly for the development of mathematical knowledge.

The Significant Bilingual Instructional Features Study (Tikunoff, 1985, no date) is the only bilingual-education-program study that I have found to specifically investigate the quality of instruction that LEP students received not just in terms of English-language development, but also in terms of academic development. Tikunoff and his colleagues used models of direct instruction to assess the quality of instruction in bilingual classrooms where native language instruction was in Spanish, Chinese, or Navajo. Unfortunately, their study design commingled mathematics instruction with instruction for other subjects, and it also used time-on-task as a proxy for achievement. Tikunoff's (no date) description of effective instruction in bilingual classrooms is very consistent with -- though not as highly structured as -- Active Mathematics Teaching (Good, Grouws, & Ebmeier, 1983). Beyond direct instruction, Tikunoff and his colleagues identified three teacher behaviors that mediated the effectiveness of direct instruction for LEP students. Effective teachers in Tikunoff's study used both English (L2) and the NES/LES students' native language (L1) for instruction (p. 12). They focused on developing NES/LES students' language, both L1 and L2 (p. 13). And, they responded to and used cultural information during instruction (p. 14). Lending weight to Tikunoff's findings is the fact that direct instruction also has been identified as a characteristic of effective instruction in Chapter 1 settings (Kennedy, Birman, & Demaline, 1986).

Summary Comments

Bilingual-education-program research and evaluation have been driven by concerns for the development of English and of academics among LEP students. These studies have taken for granted the school mathematics curriculum that LEP students are exposed to and, even when problems in instruction are noted, those concerns get cast in terms of language development.

On the one hand, by accepting curriculum and instruction as programmatic given, it has been possible to design and implement evaluations and research studies of increasing sophistication. We really have learned a few things about mathematics teaching for LEP students over the past years, and it would be foolish to pretend that we haven't. It might seem tempting to conclude that we really should continue with business as usual. What is needed, one might be tempted to say, are some better studies that seek to merge mathematics with English-language curricula or that try to document how instruction in mathematics might support the development of language skills. To these efforts, one might recommend adding some attention to closing the achievement gaps, to long-term goals such as advanced coursetaking, and to out-of-school and social goals. But by and large, it might be tempting to not change in any fundamental ways current practices in bilingual-education-program evaluation and research. In the following section, I will argue against such a position. That argument is based on the fact that the general school mathematics curriculum and its teaching have been found wanting on a variety of grounds.

The Shifting Target

Let us assume for a moment that we were able to achieve some of the goals outlined earlier in this paper. Assume that we could close the mathematics-achievement gap between LEP students and their English-proficient peers. Assume further that the gap would remain closed and that these students would enroll in mathematics courses in numbers that were comparable to those of their peers. Though this would be quite an accomplishment, should we be happy with it? If we are to believe the plethora of reports that have come out over the past years, the answer is a resounding NO (American Association for the Advancement of Science [AAAS], 1989; American Mathematical Society [AMS], 1990; Mathematical Association of America [MAA], 1989, 1990, 1991; Mathematical Sciences Education Board [MSEB], 1990; National Council of Teachers of Mathematics [NCTM], 1989, 1991; National Research Council [NRC] 1989, 1991; Steen, 1990). In the event of such success, all that would have been accomplished is that LEP students would be performing at levels that are judged inadequate when compared to international standards (McKnight, Crosswhite, Dossey, Kifer, Swaford, Travers, & Cooney, 1987; Stevenson, Lummis, Lee, & Stigler, 1990; Stigler, Lee, & Stevenson, 1990). In addition, today's students are encountering insufficient amounts and the wrong kinds of mathematics for what they will need to participate meaningfully in the United States' democratic institutions, in a changing worldwide economic order and its social systems, in the workplace, and for purposes of national security (Secada, 1990b, 1991b; Zarinnia & Romberg, 1987).

The Agenda for Action (NCTM, 1979) argued that problem solving and not the development of basic computational skills should be the focus of school mathematics instruction. Since that time, consensus has been building within the mathematics education community -- comprised of researchers, practitioners, supervisors, and other interested publics -- on a new vision for the content and teaching of school mathematics (Romberg, in press; Romberg & Stewart, 1987). That consensus has been articulated in a series of documents that lay out an agenda for reforming school mathematics in the United States. That agenda is focused on the development of new goals for school mathematics and the development of curriculum, teaching, student assessment, and program evaluation that can support the attainment of these new goals (NCTM, 1989, 1991).

New Goals for School Mathematics

According to the National Council of Teachers of Mathematics (1989), the overarching goal for school mathematics should be the development of a mathematically literate society. For individual students this means the development of

> mathematical power...[or] an individual's abilities to explore, conjecture, and reason logically, as well as the ability to use a variety of mathematical methods effectively to solve non-routine problems....Mathematics [is] more than a collection of concepts and skills to be mastered; it includes methods of investigating and reasoning, means for communication, and notions of context. In addition,...mathematical power involves the development of personal self-confidence. (p.5)

Specifically, the NCTM has proposed five goals for school mathematics. Each student should (1) learn to value mathematics; (2) become[5] confident in her or his abilities to do mathematics, (3) become a mathematical problem solver, (4) learn to communicate mathematically, and (5) learn to reason mathematically (NCTM, 1989, pp. 5-6).

Other writers have approached the specification of more general curricular goals from a different perspective than that of NCTM.

Archbald and Newmann (1988) have written about authentic achievement as that which involves the use of disciplined inquiry to produce knowledge (and a product) that has personal, aesthetic, or social value beyond completing the procedures of school. For authentic achievement in mathematics, goals and school tasks would have to be specified so as to have the aforementioned values that would link mathematics to the world outside of school.

Student Thinking in Mathematics

If curricular goals represent the targets for educational practice, then student thinking is the starting point. The most common criticism of current practice in curricular materials, the content of course coverage, instruction, and assessment is the chasm between how people actually think and learn versus how children are expected to learn in school mathematics. For example, Carpenter (1985) has argued that children enter primary school with much more competence in mathematical reasoning than they are credited with. But, the first-grade arithmetic curriculum, with its stress on memorization of basic facts rather than on problem solving, ignores that competence, and thereby, it lays the groundwork for future school failure. In a later paper, Romberg and Carpenter (1986) built a similar case in criticizing direct instruction for ignoring student thinking.

Similar to writers from within the mathematics education reform movement, Resnick (1987a) has argued that one of the primary functions for schools is teaching students to learn to think. But while writers from mathematics education have chosen examples that are clearly connected to the discipline, Resnick diverges somewhat by drawing on how people learn outside of school (Resnick, 1987b). For example, she describes how knowledge is accumulated and distributed within complex organizations, such as on a large boat, and how individuals have but a portion of the knowledge that is required for the organization to function properly (Resnick, 1987b). Examples like these are more closely aligned with Archbald and Newmann's (1988) notions of authentic learning than the more discipline-based examples found in the NCTM (1989) Curriculum and Evaluation Standards. These different nuances in meaning have implications for teaching and assessment; more on those points later.

Regardless of the disciplinary content of student thinking, there seems to be broad consensus about the nature of that thinking and of learning. Thinking, problem solving, and to some extent learning are thought to share similar characteristics of sense making and of relating new information to established knowledge. Where disagreements occur is in interpretation of the specifics. Information processing models of thinking, for example, require detailed specifications of conditions and of productions that occur under those conditions (e.g., Seigler, 1991). The anthropological study of how knowledge is produced, on the other hand, focuses on practices within cultural groups that are thought to create that knowledge and on the social processes by which that knowledge gets validated (e.g., Lave, 1988).

According to information processing and cognitive science theories, knowledge develops in one of three ways: through the gradual accretion of new information to what is already known, through the exposition and resolution of areas of conflict, and through the reorganization of existing knowledge structures. Within the more anthropological traditions, knowledge is thought of as an artifact of human activity. It derives its meaning and validation from that activity and how the activity gets situated within the larger social setting. Hence the processes of knowledge acquisition must be linked to the contexts in which people produce that knowledge.

Many researchers in mathematics education have characterized knowledge as consisting of conceptual and procedural parts (Hiebert, 1986). Conceptual knowledge is interconnected and rich in relations; procedural knowledge produces something. This distinction can be thought of as roughly parallel to the distinction between number concepts (e.g., knowing the concept of 5) and the ability to compute (e.g., knowing how to obtain 2+5). According to Hiebert (1986), mathematics teaching should help students develop and link both sorts of knowledge.

Alternatively, writers who are grounded in information processing models of thinking tend to posit the existence of three broad categories of knowledge: conceptual, procedural, and also strategic (Siegler, 1991). Roughly speaking, one can think of an information processing system as composed of its production rules (procedural knowledge), the conditions that must be met for the system to operate (conceptual knowledge), and an overarching operating system that monitors and regulates the entire process from beginning to end (strategic knowledge). Problem solving consists of the orchestration of all three sorts of knowledge to attain a goal.

Thus even within similar cognition-based approaches to the study of student thinking, there are subtle differences. These differences get played out in different approaches to the specification of curricular tasks, to tracking, and to assessment.

Curricular Tasks and Instruction in Mathematics

If the goals specify the end points, and if student thinking provides the beginnings as well as constraints for school mathematics, then curricular tasks and instruction should provide the means by which to develop student reasoning and thinking to the desired end points. Again, there is a broad consensus that tasks and instruction should be aligned to the new goals and that they should support the development of student thinking.

Mathematics Curriculum and Tasks

The curriculum has been faulted for failing to produce desired outcomes, for being a disconnected hodgepodge of content, and for lending itself so easily to superficial coverage (Freeman & Porter, 1989; Porter, 1989; Porter, Floden, Freeman, Schmidt, & Schwille, 1988). This lack of cohesion and superficiality do not support the development of conceptual knowledge or of links between conceptual and procedural knowledge (Hiebert, 1986; Romberg & Tufte, 1987). Moreover, this content fails to provide students the disciplinary experiences that they need to develop mathematical power (NCTM, 1989) or the authentic tasks that are necessary for authentic learning to take place (Archbald & Newmann, 1988; Resnick, 1987b).

Hence, new tasks should be developed and organized to provide greater coherence and more depth of coverage (Archbald & Newmann, 1988; Romberg & Tufte, 1987). Those tasks should reflect disciplinary forms as well as authentic forms of mathematical knowledge (Archbald & Newmann, 1988). They should provide students with opportunities to solve problems, to reason mathematically by making conjectures that are then socially validated, to communicate with one another using mathematical language, and to make connections among a variety of representations of the same problem situation (NCTM, 1989). Paper-and-pencil computational facility should be deemphasized; i.e., things like arithmetic algorithms and the solution of algebraic equations through the manipulation of written symbols should be relegated to calculating devices such as calculators and computer software. In place of computations, discrete mathematics, geometry, linear programming, measurement, probability, statistics, and other content should be emphasized (NCTM, 1989).

Some mathematicians go even further in their recommendations for reorganizing the school mathematics curriculum. Steen (1991) and his collaborators would organize mathematics around common themes, like the study of patterns, that cut across and unify seemingly disparate mathematical fields like geometry and statistics. Alternatively, Kaput (1991) has argued for totally scrapping the high school mathematics sequence of Algebra, Geometry, Algebra II, Trigonometry. In its place should be a unified-mathematics sequence that includes new content; relegates all symbolic manipulations to calculators, computers, and other technologies; and uses these technologies to develop depth of understandings and relationships among the different fields of mathematics.

In spite of this agreement on broad goals, there is an emerging tension between the disciplinary and psychological goals of developing mathematical power (NCTM, 1989) versus the criterion that authentic tasks should have external personal, aesthetic, or social value (Archbald & Newmann, 1988). Many tasks found in the mathematics reform documents, while having great disciplinary value, seem to have very little value outside of school. Many tasks that seem very authentic cannot be accomplished within the constraints of the school term, but what is more problematic from a disciplinary point of view, they can be done without reliance on deep mathematical principles.[6]

If mathematics is to be undertaken within rich, real-world problem settings, then another area for debate emerges around the settings that will be chosen for study and therefore will be granted legitimacy as worthy of mathematical scrutiny (Frankenstein, 1989, 1990; Secada, 1991b; Stanic, 1991). In part, this debate revolves around questions of whose interests are served by the study of those contexts and how students are socialized through that study, either explicitly or tacitly (Secada, 1991b). For example, adult students in Frankenstein's (1990) intermediate algebra class learn about percentages by studying how decreasing rates for electricity are linked to increased consumption, and that increased consumption most often entails using appliances that only the wealthy can afford (air conditioners, pool filtration systems, and the like). This analysis of consumption is based on social class. It is in sharp contrast to a mathematical analysis wherein decreasing rates for increased consumption are made to seem as the natural and inevitable outcomes of the so-called laws of supply and demand.

The study of mathematics through authentic contexts also socializes students into accepting certain norms of behavior. For example, a very common activity in elementary school is for students to operate a store of some sort. What seldom, if ever, occurs is for students to run a social-service agency that provides services either for free or on a sliding scale. Presumably one could develop and study exactly the same sorts of mathematical knowledge and skills in either context; yet one context gains legitimacy, the other does not.

Thus, while there is broad-based consensus that mathematics tasks need revamping to support the development of student reasoning, there remain questions about (1) how the new tasks will be organized; (2) the tension between disciplinary knowledge and authenticity; and (3) the cultural contexts that get represented in the curriculum and that thereby will receive legitimacy as being worthy of mathematical study.

Mathematics Instruction

Again there is broad consensus that instruction should support the development of student reasoning, communication, and similar processes that are thought to enhance student learning (Hiebert, in press; Idol & Jones, 1991; Jones & Idol, 1990; Lampert, 1988, 1990a, 1990b; NCTM, 1991). There are some debates about whether or not direct instruction -- as it has been classically understood -- can support student reasoning, especially among students in compensatory programs like Chapter 1 (Brophy, 1991; Collins, 1991; Collins, Hawkins, & Carver, 1991; Idol, Jones, & Mayer, 1991). Some writers who have grounded their analyses of student reasoning from an information processing point of view have argued that direct instruction can incorporate the teaching of specific thinking skills (Idol, Jones, & Mayer, 1991) or that it can include cognitive supports that are slowly withdrawn as students take on increasing responsibility for their own learning (Collins, 1991; Collins et al., 1991). Yet these analyses remove or transform many of direct instruction's defining characteristics -- for example, teachers would no longer directly tell students what they were to learn. Thus, it is not clear that direct instruction as it has been classically understood remains a viable instructional strategy.

Others writers are arguing for a radical overhaul in what constitutes good teaching of mathematics (Ball, 1990; Lampert, 1988, 1990a, 1990b; NCTM, 1991). According to them, teaching is a question of orchestrating student engagement in worthwhile mathematical tasks. A teacher does not tell, but rather he or she poses problems and organizes students into groups to work on those problems. The teacher provides social supports for problem solving, challenges students to justify their responses, and helps students to amplify their justifications when those justifications are not fully developed. The teacher establishes norms of behavior wherein students are to be comfortable participating and are to allow and encourage others to contribute, even when those contributions later do not survive public scrutiny by the whole class.

There are approaches that seem to lie between direct instruction and these more radical departures and they may include features from both. For example, Japanese and other Asian teachers are thought to teach mathematics by spending most of their time on lesson development in whole-class lecture settings. They support student reasoning by discussing one or two problems in great depth, trying to solve them in as many ways as possible. Also, they orchestrate classroom discussion around each student's strategies and try to expose misconceptions as opportunities to revisit and reteach important ideas (Stigler & Stevenson, 1991).

Another approach, known as Cognitively Guided Instruction (CGI), (Carpenter & Fennema, in press; Carpenter et al., 1990; Peterson et al., 1991), combines insights from over five decades of research on how children solve addition and subtraction word problems (Brownell, 1928; Carpenter & Moser, 1984) with more recent research on teacher decision making (Clark & Peterson, 1986). CGI is based on four interlocking assumptions: (1) teachers should know how mathematical content is organized in their children's minds; (2) teachers should make mathematical problem solving the focus of their instruction; (3) teachers should find out what their students are thinking about the content in question; and (4) teachers should make instructional decisions (e.g., the sequencing of topics) based on their knowledge of their students' thinking. Unlike other programs and approaches that prescribe teacher behaviors, this approach relies heavily on teachers' basing their instructional decisions on their knowledge of how their students are thinking about the content (tasks) that they are engaged in.

There are many other issues for instruction, among them, classroom organization. Should the whole class participate in an activity, should it be small groups, or should it be individually based? If instruction is organized by groups, should they be by mathematics ability or heterogeneous? Since mathematics is a social activity, social interaction is necessary. Such interactions are possible not only with small groups but also in whole class settings (e.g., Lampert, 1988, 1990a, 1990b).

Assessment and Evaluation in Mathematics

Goals provide the end point; student cognition, the beginnings and the focus of teaching; and tasks and instruction, the means for achieving those goals. Assessment and evaluation provide evidence that the goals are being met. Assessment focuses on the student; evaluation, on the overall mathematics program in which the student is enrolled.

The NCTM (1989) Curriculum and Evaluation Standards outlined eight aspects of assessment and evaluation, each composed of two poles. One pole should receive emphasis, the other should be deemphasized (p. 191). For example, while assessment should focus on what students know and can do, decreased attention should be placed on what students do not know. Assessment should be ongoing and integral to instruction, not solely for the purpose of assigning grades. And in program evaluation, standardized achievement tests should be one of many possible indicators for monitoring success; other indicators should include samples of student work that are collected in a variety of settings and through a variety of methods. Beyond agreement on general principles like these, however, there remain points of debate within fields of both assessment and evaluation.

Assessment

Most simply, argues the NCTM (1989), assessment should be aligned to the new curricula that are intended to achieve newly developing mathematical goals. Indeed, one of the most common complaints about current practice is the failure of tests to be properly aligned to curricula that are in place, even today, and the total misalignment between curricula of the future and present-day tests.

From these misalignments have come two major hypotheses. First is the hypothesis that tests are determining what students actually encounter in their classrooms even if there are broader curricular objectives than are measured by the test (Romberg, Zarinnia, & Williams, 1989; Resnick & Resnick, 1991; Silver, in press). If tests actually are such strong determinants of what gets taught to students (for a counter argument, see Porter, 1989), then current testing practice will derail efforts to reform school mathematics. However, if tests really are such strong determinants of curriculum, then an alternative becomes available; by changing the test, we can change what gets taught (Silver, in press). If we change the tests to include tasks and items that approximate emerging goals for school mathematics, curriculum and instruction will follow. California and Connecticut have adopted this strategy; the former includes open-ended items in its assessment and the latter is committed to using only authentic assessment.

Silver (in press) has argued that this hypothesis may be overly optimistic. Teachers might teach based on a variety of things, not just what is tested -- for example, how they were taught, their beliefs about what constitutes "real" mathematical knowledge, or the press to cover the book. Thus, tests and the curriculum that students are exposed to may be determined by similar forces, but testing per se does not determine the curriculum. Efforts to change curriculum by changing the tests will fail, if not backfire, because they would not address the deeper causes of why teachers teach as they do.

A second hypothesis growing out of the misalignment between testing practice and future curricular needs is a weaker version of the first. Tests, at least standardized achievement tests, are but one of many indicators that teachers rely on in their practice. Because results are so seldom returned to teachers quickly enough or in a format that enables them to make instructional decisions based on the results, standardized tests are, ultimately, unimportant vis-à-vis practice. Their importance lies in their symbolic value, as indicating a job well done or providing the public with evidence that the schools are working. Teachers attend to tests not because it will help their practice but because it must be done to placate outside interests.

Tests may fail to reveal what students are actually doing in their schools and may interfere with their instruction not because they are dictating curriculum but because they are add-on nuisances.

Paper-and-pencil standardized tests have little utility. As part of an overall reform effort, they need to be changed to support (or at least, not to interfere with) curriculum reform. Be it to mandate or to support curriculum reform, there is consensus: assessment in school mathematics needs revamping (also see Resnick & Resnick, 1991).

Regardless of this consensus, there are still many issues about mathematics assessment that must be worked out. These issues include debates about the kinds of tasks that will comprise these new assessments, the conditions under which they will be completed, what work must be exhibited, scoring rubrics, the creation of performance standards, how to communicate the new rules of testing to participants, and how to communicate the results so that they are meaningful and useful (also see Lajoie, 1991).

Assessment tasks. Beyond agreement that new assessment tasks need developing, there are few exemplars of such tasks and fewer still that would meet the range of criteria found in the various reform documents. An item from the Connecticut assessment reminds students how to compute the volumes of a sphere and of a cone. The task provides a context wherein a scoop of Ben and Jerry's ice cream is placed on a wafer cone. The ice cream forms a perfect sphere of a given diameter. The wafer cone forms a perfect cone of given diameter across the base, with equilateral sides, and is of a given height. The problem is to determine whether the cone could hold all of the ice cream were it to melt.

The samples of some students' work coming from this task are impressive. They clearly understood the need to delve into the mathematical properties of the task for purposes of this assessment. However, this task fails to meet criteria for authenticity as outlined by Archbald and Newmann (1988) or Resnick and Resnick (1991). How could anyone produce a perfect sphere and why would anyone allow a scoop of Ben and Jerry's ice cream to melt -- unless it was for school?[7]

Nominally authentic tasks may fail to reveal the types of mathematical reasoning that are called for in the various documents. For example, students enrolled in an alternative high school conducted surveys of their peers on various topics. They designed, distributed, and collected the surveys. The students then compiled the results and reported them to the entire school by displaying the results of each month's surveys on the school's bulletin board.[8] On the one hand, it would be very easy (and very authentic) to enter the results of each survey into a software program that would compile them and generate appropriate charts and graphs for display. Such an approach also would close off any opportunity for students to develop and display mathematical competence of the sort that is called for in the reform documents. However, the teacher's intent for this task was to develop mathematical reasoning involving parts of whole, percentages, and the graphical display of data. He did not use the school's readily available computer lab in this activity. Instead, students compiled the results by hand. They converted the results for each question into percentages using calculators, and then they displayed those percentages in pie charts using compass and protractor. In other words, this teacher sacrificed some authenticity in order to develop and to assess student mathematical reasoning.

Authentic assessment tasks are open to the same questions about the standards by which their authenticity is judged as are curricular tasks. For example, the California Assessment (Stenmark, 1989) includes a task wherein students are told that a local college accepts one half of that high-school's graduating class each year while another college also accepts one half of that graduating class. An individual student believes that he is certain to be accepted to one of these two local colleges. The problem is to explain what is wrong with this student's reasoning. At first glance, this task has much out-of-school value, until one realizes that over half of all graduating seniors do not go on to college. One must ask if non-college-intending students would have more than minimal interest in such a task. It seems unlikely that this task will reveal what uninterested students really can do.[9]

One possibility for overcoming problems about cultural and other forms of bias is to allow students to choose among many tasks that include a broad range of cultural contexts, and require comparable mathematical thought, and are to be finished within similar time constraints. For example, a student might enter two raffles for the right to purchase tickets to over-subscribed rock concerts. For the first concert, the odds of winning a pair of tickets would be 50-50; and for the second, the odds could be 60-40. Would this student be assured of getting in to see one of the two shows? Including additional settings increases the likelihood that students will be sufficiently intrigued by at least one of them to actually apply themselves to the task. Some students may actually see the structural parallels among such tasks.

In the past, the search for unbiased test items has meant a search for items that could cut across social class, gender, and cultural categories. One reason that we have such impoverished curricula and tests may be the difficulty in creating such "neutral" tasks. A better strategy might be to create many tasks representing a range of cultural contexts and to ask students to pick the ones that intrigue them the most.

Conditions for assessment. How much time should students have to produce their work? State assessments typically last one to two hours, so that tasks for these assessments would have to be finished within some rather tight time limits.

Such time limits, however, would fail to demonstrate what students could do when engaged in long-term projects. For example, one prototype task developed at the Center on Organization and Restructuring of Schools (CORS)[10] for tenth-grade students provides a setting wherein a family of four moves from Madison, Wisconsin, to the city where the students who are engaged in this task live. The students are given the Thursday and Sunday newspapers of both of these cities since those issues contain information about homes and apartments (for rental or purchase), food, clothing, different kinds of sales, entertainment, job opportunities, and the like. The students are told that, in order to live in Madison, this family spends a certain amount per month that is allocated among these and other budget categories in a certain way. The first problem for the students to solve is: In order to maintain a comparable standard of living, how much per month will this family of four need to spend? Secondly, the students are told to assume that a different group's estimate is double theirs; How would they convince that group that theirs was the right estimate? Third, the students are told that, in order to have so much disposable income, people must earn more since they must pay taxes, social security, health and medical benefits, etc. If in Madison, this family of four's take-home pay was based on a given earned income, what would the earned income have to be in their new home town? And finally, assuming that two people worked in this family, What sorts of jobs would they have to have in order to make ends meet in their new home town? A task like this simply cannot be done in two hours.

Should assessment tasks be uniformly created and administered by an outside agency? Should they be samples of student work that are collected over the course of the year and represent a common core of important tasks as identified by the teacher? Or, should students select their best work and place it into a portfolio that then gets graded? Under current notions of accountability, the first option would be desirable. When issues of curricular validity and alignment are foremost (or when teachers will be evaluated based on their students' work), the second option would seem preferable. However, if one is strictly following models of authenticity -- i.e., what real people do in the real world -- then the last option would be preferred.

Actors, architects, artists, musicians, and even professors up for tenure and promotion assemble their best work for review.

What work should actually be displayed? Testing programs like the California Assessment, the Connecticut Assessment, and many curriculum development projects ask students to show their work enroute to achieving their solutions. This is because these assessments are looking for evidence of mathematical reasoning, communication, and the creation of new knowledge. However, according to standards of authenticity, what should be required are samples of finished work, not the work that was produced while the finished product was being developed. An architect does not include sketches and initial renderings in the final product; musicians do not include rehearsal tapes in their portfolios; business people do not give all the details of why they recommend something in their memos; nor do mathematicians include the false starts in their final articles.

The difficulty with asking for a final product, however, is that it hides the disciplinary work that went into its production. Consider, for example, the task described earlier wherein high school students surveyed their peers and presented the results of those surveys to the school. It is difficult, if not impossible, to determine what mathematical understandings these students actually used in creating those pie charts. For example, we might assume that in the production of these pie charts, students had to have converted individual responses into percentages; i.e., for question number 4, the number of students responding a, b, c, or d would have to be converted into the percent of the students who chose each of these options. We might assume that, in making this conversion, a student had demonstrated knowledge of how parts of a whole are related to percentages.

But consider the case of the student that I observed working on this step of the task. Someone else had already converted the raw scores into percents for all 20 questions on the survey. But she had noticed that the percents for each question did not always add to up to 100. When she pointed this out to her teacher, he told her that not every student had answered every question; for example, 30 of the 31 surveys that had been returned included a response to question 2. Thus, this student was busy checking all of the questions; for those that did not add up to 100 percent, she would divide all of the responses by 30, i.e., by how many people had answered question 2 and not by how many people had actually answered each specific question.

After recomputing the percentages for the questions that needed to be recomputed, she checked her totals and became very distressed when many of them still did not add up to 100! I asked her if she knew why they should add to 100. She responded because she had learned it in another class. Then, I pointed out that, in some cases, her sums came out to 99 percent and she didn't seem to mind that, but that when the sums came out to 98 percent she got upset. She did not answer that there might have been a rounding error or even that 99 was closer to 100 than 98. Instead, she commented that 99 was just a better answer to get. So next I asked her why she was dividing everything by 30. Her answer was that she had done so for question 2 and in that case the percentages added up to 100.[11]

Since she did not link these questions about parts and whole to how she might resolve her dilemma, I tried to explain to her that a different number of people had responded to every question on the survey. For instance, 30 people had responded to question 2, but only 29 had responded to question 4 and, of those 29, 11 had chosen a. She still did not make the connection that she needed to divide the 11 by 29 because 29 was the appropriate whole for that particular question on the survey. What percent of 29 is 11? I continued. Still no response or indication of understanding. Instead, she kept insisting that the answers had to add up to 100 percent. She did not see how, by dividing the responses for question 4 by 29, she would satisfy this condition. Finally, I suggested that she simply try doing so. Afterwards, I suggested that she add up the percents to this question one more time. Of course, they totaled 100 percent; but, rather than try to understand why this particular example worked, she adduced a general rule -- divide the response to each question by the number of people who actually responded to that question. And very happily, she proceeded to complete the task.

This episode raises many issues in terms of how this student was linking (or failing to link) conceptual understandings about percentages and parts of a whole to the procedures by which she was converting individual student responses to aggregate percentages. However, in this and every other student's final product, there will be no evidence about whether or not such understandings were created, strengthened, or even used. All that remains are the end products of that effort. It is not surprising, therefore, that students are told to show their work in an effort to determine whether they are displaying the forms of mathematical reasoning that the tasks are meant to support.

On the other hand, I have seen samples of work where students were scored lower because the work that they displayed lacked coherence, which is exactly how work in progress is characterized. In one extreme case, I saw a short, concise explanation wherein a student had gone straight to the heart of the task and had done so elegantly. But this work was in a lower corner of the page, lost in the jumble of his other work. We are still struggling to find some middle ground on this issue.

Scoring rubrics. As noted earlier in this section, one of the reasons students are asked to show their work is in order for someone else to score the quality of that work against certain standards. The actual content of those standards is still under discussion. Task performance could be scored according to a learning theory or some other criterion (Lajoie, 1991).

In CGI, first grade teachers are taught to assess their students' knowledge on an ongoing basis (Carpenter & Fennema, in press; Carpenter et al., 1990). In those assessments, teachers rely on a well structured body of research on learning wherein a student's right or wrong answers can be linked to how difficult that problem was either in terms of its semantic structures or in terms of the size of the numbers that were used. The strategies that children use when solving various word problems also provide teachers with information about how their students are thinking of the problems. While not written down as formal scoring rubrics, these assessment techniques rely on judgments that are linked to a very rich and detailed specification of how children learn, i.e., to a highly localized learning theory.

Where such specificity is not possible -- in most of the rest of school mathematics -- we could still generate scoring rubrics based on more general learning theories. For example, cognitive scientists (Siegler, 1991) often posit the existence of three kinds of knowledge: conceptual, procedural, and strategic. Lane (1991) included these categories of knowledge in scoring rubrics that were developed for assessing middle school students' performance on a range of authentic tasks.

Alternatively, non-psychological criteria could be developed. Stenmark (1989) describes a general scoring rubric that was used in scoring the open-ended questions of the California Assessment. This rubric was used to score student performance on open-ended questions based on the clarity and coherence of the response; the appropriate use of pictures or diagrams; the quality of the presentation to the intended audience; the use of mathematical reasoning, ideas, and processes; and the nature and flow of the argument that was developed in the response.

If one follows the tenets of CGI and if assessment is supposed to serve instructional purposes, then scoring rubrics should combine explicit learning theories for the tasks at hand with some way of targeting that learning to a coherent end point. While pedagogically these would be the most useful rubrics to develop, they also are the most difficult. We simply do not have as detailed models for how students learn mathematics in domains outside of arithmetic word problems. More general rubrics, like that developed by Lane (1991), may be the best that we can do. The utility of such rubrics for instructional or accountability purposes would remain an open question.

Alternatively, one could create scoring rubrics based on out-of-school models of adequate performance. In some settings, conciseness is more important than the flow of an argument. For portfolios, no learning-theory-based scoring rubric may be adequate, since portfolios are supposed to contain only finished work. Moreover, in the real world, portfolios are scored on a case-by-case basis; i.e., an individual's (or a group's) work is evaluated anew every time that person seeks employment.

Performance standards. To be used, scoring rubrics must contain not just the content or dimensions of interest but also standards against which to judge how people actually perform. How those standards should be developed and calibrated remains an open issue. For example, it is possible to create a priority standards by reference to some absolute criterion or by looking at how experts do the tasks. However, such standards may be set so high that no one had a chance of scoring at the top levels; they might be calibrated in such a way that pedagogically important distinctions got lost, or the experts (if their performance is used) might approach the task in ways that no one else would.

As an alternative, some people recommend that we gather samples of people's work and calibrate the rubrics against those standards. The objection to this, however, is that the performance criteria will end up, essentially, being set too low.

As new cohorts of students become more acclimated to new curricula, new instruction, and these new ways of assessing performance, it is likely that performance will improve. Hence, the performance standards that are settled on will need to be recalibrated every few years. In some sports (ice skating and diving), for instance, performance criteria are recalibrated after someone obtains a perfect score during a major competition.

Performance standards will also need to be linked to instructional practice and to accountability systems. If the standards are calibrated so high or so coarsely that all students cluster around a single level, then they will not be very helpful. On the other hand, if the standards are too finely calibrated, the scorers, teachers, and other consumers of the results may spend so much time trying to understand the distinctions between levels that they will have too little time to use the information for its intended purposes.

The new rules of the assessment game. Under the old rules of testing, students knew pretty much what was expected of them. They either got the answer right or wrong. In the case of teacher-made tests, students know to show enough work to get some partial credit in case the answer is wrong, but not to show so much work that, if the answer is right, they lose credit for work that is wrong or sloppy.

Given the many different ways of scoring performance according to the purposes and the theories that underlie each rubric, it is not clear how students will know exactly what is expected of them. Are they to produce a final, polished product? Should they omit a large amount of detail? Should they include their scratch work? If so, how can they distinguish that work, with all of its false starts, zig-zags, and lack of coherence, from the work that they wish to present? More generally, how does one communicate to a student that he or she is being scored on the use of conceptual knowledge, procedural knowledge, communication skills, or any of the other criteria that have been created? I have not seen anyone grapple with these questions, but they would seem increasingly important, especially for students from diverse backgrounds.

Communication of results. Assessment results are to serve a wide range of purposes. They must be communicated to teachers in ways that will help them make instructional decisions. Ideally such information would combine a description of student competence with some ways of placing that performance along some developmental path. Students, parents, and other stakeholders will also be interested in assessment results. How to report these in ways that all of these interested publics will understand and be able to use remains an open issue.

Program Evaluation

With so much emphasis on assessment, relatively little effort seems to have been placed on program evaluation. In part, because of the belief that, if we first change the assessment, the evaluation systems must change -- if no other way, at least in the sorts of information on which judgments are made.

The NCTM Curriculum and Evaluation Standards (1989) do provide some general suggestions. Evaluation should draw on a wide range of sources of information. Evaluation should focus on ensuring that all students (not just a few) are learning and developing their mathematical power. Evaluations should go beyond looking at student outcome data; they should also focus on the quality of the curriculum in terms of its coherence and content coverage, on the adequacy of materials and other resources, and on the quality of instruction that students receive.

In view of the originally stated purposes for the mathematics reform movement -- participation in the nation's various institutions by the next generation of students -- one long-term outcome should also be evaluated, i.e., do students who experience school mathematics as is recommended by the current reform movement actually participate in our society in the ways that they are expected to?

Summary Comments

The goals for teaching school mathematics have shifted radically. The agenda now is to shift practice to meet those goals. These new goals are focused on the development of students' mathematical power through attention to coherence, depth of study, communication, conjecturing, and the actual doing of mathematics in a variety of contexts that will support such disciplinary inquiry. As the MSEB (1991) so clearly summarized:

> Goals for student performance are shifting from a narrow focus on routine skills to development of broad-based mathematical power (p.5).

Though there are many issues that are still being worked on, there is broad consensus among mathematicians, curriculum developers, psychologists, practitioners, and many key public stakeholders that this shift is necessary because current practice is inadequate.

In current bilingual-education-program-evaluation practice, there are many points that are at odds with how school mathematics is shifting. If we continue to do more of the same, even if we do a better job, we may achieve our goals, but they are outdated and inadequate for purposes of preparing LEP students to participate in the world in which they will live their adult lives. Student assessment, program evaluation, and related research need to shift in order to match these evolving goals. What is more, bilingual education research needs to inform the mathematics education reform movement of what has been learned about the educational needs of LEP students. It is to these points that this manuscript now turns.

Mathematics Education and Bilingual Education: A Two-Way Conversion

On three points I am in total agreement with the current reform movement. First, student thinking and reasoning are the keys to this effort. We are in the business of teaching so that students can develop those skills. However, we still need to unpack what these notions mean vis-à-vis the bilingual learner.

Second, curriculum, instruction, assessment, and evaluation should be coherent, linked, and in support of the development of student thinking. Curriculum and instruction should focus on covering fewer things but providing for greater depth of coverage. Assessment and evaluation should be aligned with and support efforts to teach. They come after everything else, not by themselves.

And third, the goals for education should be linked to the larger society in which our students will live. In some places, the reform movement does not go far enough. We need to consider the situations in which LEP students (and indeed, increasing numbers of our students) find themselves. We should not shy away from the fact that many students live in desperate poverty and that education needs to help them deal with the realities of that as well.

It would be easy to make these general observations and to argue that some immediate and obvious implications for bilingual education grow from them. But the situation is more complex than would be implied by such a one-way conversation. Though changes in school mathematics may have implications for the education of LEP students, bilingual education has much to say to the reformers, not only about the education of LEP student but also about issues that include equity, culture, and performance assessment. Anyone who is even vaguely familiar with the history of bilingual education in this country should have a sense of deja vu when reading about some of the debates in mathematics education. There is much that bilingual education research can say to inform those debates.

Goals

An immediate, albeit not so obvious, implication of the changing goals in school mathematics is that the academic goals for bilingual education need to be reexamined. Academic achievement and advanced course taking should be revisited from the perspective of the kinds of courses that LEP students get placed into. One of the most often told stories for any reform is that people who are positioned to take advantage of it receive a disproportionate amount of the benefits from that change (Secada, 1991b, in press). It is important to monitor how LEP students are included in (or excluded from) reform in schools and districts and at the state and national levels. Beyond vague claims about excellence for all, we need to ensure that inclusion is meaningful.

Elsewhere, the author and others who are concerned about equity in education (Secada & Meyer, 1991) have argued that the mathematics reform movement has not paid adequate attention to these issues. For example, in laying out the reasons for needed reform in school mathematics, the Curriculum and Evaluation Standards (NCTM, 1989) gave the mathematics achievement of minorities as one of the reasons for needing reforms, yet nowhere else within the document does one find specific attention to ensuring that the proposed changes will, in fact, be helpful to minorities. To be fair, in the Teaching Standards (NCTM, 1991) there is a bit (but not that much) more attention paid to equity.

The point may seem like a subtle one. Someone could argue that student diversity need not receive specific and ongoing attention. Absent evidence that LEP students will be omitted or ill served by the reform, such efforts are covered under the rubric of reform for everyone. My counter argument is that silence on issues of student diversity leaves open the very real possibility that, within the reform of school mathematics, stratification of students along the lines of race, social class, language proficiency, or some other means will be recreated. For example, one of the people whose practice is held up as an exemplar for mathematics teaching is Magdalene Lampert (1988, 1990a, 1990b). Anyone who reads her thoughtful analyses of teaching and the skillful ways by which she focuses on student understanding should be impressed by the vision of teaching and the possibilities that she describes. In a recent paper, Lampert (1990a) wrote about her efforts to construct meanings for fractions and computations in her fifth grade classroom. In one particular vignette, she discussed how a community of discourse was formed and maintained in the class.

> Students asserted their contributions and other students revised them. The end result was produced with little teacher input, except asking for clarification and recording on chalk board what was said. All but four members of the class made an active contribution to this discussion; two of the students who did not contribute had very limited English-speaking ability. (p. 263)

In other words, half of the students who were omitted from the community of discourse for this episode were limited English proficient.[12] Note how it seems as if these students' limited English proficiency is the reason for their nonparticipation. Thereby, their exclusion from a discourse community (which is by definition a social fabrication) is made to seem natural and is legitimated.

And this is precisely my point. By their failure to specifically include equity and student diversity as concerns that are integrated from the very start, the various reform documents make possible the restratification of opportunity along the lines by which it has taken place in the past. As the goals get articulated, we must continually ask, Who are the goals for? People in bilingual education must advocate meaningful inclusion of LEP students.

The long-term and out-of-school goals for bilingual education need revisiting. One of the reform movement's main pillars is that today's students need preparation for tomorrow's world -- including access to jobs and meaningful participation in our society. Such goals are commonly missing from similar discussions in bilingual education on grounds that the development of English is the more pressing concern. It would be a major contradiction, however, to argue that the goals for the mathematics education of LEP students should be linked to out-of-school outcomes but that the goals for the larger program should not.

As part of the concern for social goals, we need to move somewhat beyond the development of mathematical power. We must ask about the knowledge and skills LEP students must have in order to participate meaningfully in American society. There is ample research -- much of it being carried out by people involved in bilingual education -- to document how people are discriminated against due to skin color, accent, and the like. Recently, for example, the Secretary of Labor issued a report that documented the glass ceilings that women and minorities encounter in large U.S. corporations. The question cannot be avoided: What must LEP students know and be able to do in order to overcome those barriers? The answer is likely to include much of what the reform documents say, but it is also likely to diverge in some significant ways.

This general issue is also one that the mathematics reform movement needs to address. Everybody Counts as the NRC (1989) avers. But the question remains, Counts for what purposes? Answers from the bilingual education community should inform a similar debate in the mathematics education community.

The Bilingual Learner of Mathematics

We need to create a view of the LEP student as a learner of mathematics that combines what we know about how mathematics is learned with what we know about second-language learning. Among current efforts that could be helpful in this regard are research on learning strategies (O'Malley & Chamot, 1990), content-based ESL (Crandall, 1987), and the relationship between bilingualism and enhanced functioning in the academic areas (Hakuta, 1986; Secada, 1991a).

It may be helpful to look for common learning processes that cut across language learning and mathematics learning. Such domains might include psychological processes that are common to understanding language and mathematics (Kintsch & Greeno, 1985) as well as for producing either linguistic or mathematical output once someone understands something; sociolinguistic and cultural processes that support the creation of discourse communities in school and how sensemaking takes place and gets validated within such communities (Heath, 1986; Lampert, 1988, 1990a, 1990b; Lave, 1988; NCTM, 1991; Simich-Dudgeon, McCreedy, & Schleppegrell, 1988/89); and how variation in sociocultural contexts affects performance (Stanic, 1991; Zentella, 1981). Of course, distinctions based on content will need to be made; obviously, the retelling or translating of an arithmetic word problem calls on, at some point, different processes than the solution of that problem.

We need to be careful that the analyses of how bilingual people learn mathematics are not always seen as derivative of research employing monolingual populations. Many analyses are based on the notion that bilingual people are the minority and that research concerning them can be thought of as an application of what we have learned about the majority. This assumption, however, is simply wrong. The norm, within the world, is to be bilingual (Skutnabb-Kangas, 1988).

The research issue is not just the adapting of research concerning monolingual populations to bilingual populations. The more basic research issue concerns the generalizability of results that were found in monolingual populations to the case for bilingual ones. It may well be that much research does generalize. But we cannot tell since we have not developed a unified view of the bilingual learner of mathematics. In a real sense, we are only beginning to learn how sense making occurs in such populations and hence what it means to say that student reasoning -- for the bilingual student -- is the starting point for school mathematics. There is much work to be done.

Curriculum and Instruction

The simplest and most straightforward implication of the mathematics reform movement to the case for bilingual education is that curriculum and teaching for bilingual learners should support the development of their mathematical reasoning. But since we are not clear on the full scope of such a claim, much work still remains in the area of curriculum and instruction.

One promising line of work might be to expand notions that have been found in content-based-ESL and language-learning approaches to create a more unified view of the tasks and instructional methods. It would be helpful to understand where structural analyses of what has become known as the mathematics register (Crandall et al., in press; Dale & Cuevas, 1987; Spanos et al., 1988) diverge from sociolinguistic analyses of communication in classrooms, specifically in mathematics classrooms (e.g., Cazden, 1986; Lampert, 1988, 1990a, 1990b). In the structural analyses, meaning seems somehow to reside in the language and symbols of mathematics. Not surprisingly, direct instruction is used to develop such meanings (e.g., Chamot & O'Malley, 1988). Alternatively, sociolinguistic analyses are more dynamic. They place the development of meaning for symbols within contexts where those symbols are needed to communicate mathematics in meaningful and unambiguous ways -- much as whole language approaches to reading place the development of vocabulary in context.

There may be more value, from the standpoint of curriculum and teaching, in relying on sociolinguistic as opposed to structural analyses of the mathematics register. Such an analysis would seem more consistent with how Cazden (1986) describes a register as a sociolinguistic construct. Structural analyses of the mathematics register may also increase the fragmentation in the mathematics curriculum for LEP students. Not only are there lessons for skills development but also for mathematics vocabulary and symbolism. This does not mean that structural analyses of mathematical language are not helpful. Indeed, the addition and subtraction problem solving literature relies very heavily on them (Carpenter & Moser, 1984; Secada, 1991a). But a more unified approach would seem, at present, to be called for. It may turn out that attention to higher level structural units -- such as paragraphs, texts, and discourse frames -- will provide greater payoffs than in the past.

The debates on social and cultural referents in mathematics tasks could be informed by similar debates within bilingual education. If mathematics educators are going to take seriously questions of out-of-school outcomes and task authenticity, then they also will need to attend to the situations in which bilingual learners live. Work by Moll, Velez-Ibanez, and Greenberg (1990) in literacy development might provide some ways of proceeding here. A range of social and cultural contexts will need to be represented in newly developing mathematical tasks. We need to develop guidelines for including contexts that are unfamiliar to mainstream cultures and ways for teachers to capitalize on the mathematics that can be learned in such settings.

Newly developing models for teaching mathematics should be scrutinized for their applicability to bilingual learners and adapted as necessary. Lampert's (1990a) acknowledgement of the limitations in her teaching is a reason to question but it is not a reason to reject the developing visions for teaching mathematics (NCTM, 1991). Maybe, with some adjustments -- specifically inviting these students to add their thoughts, encouraging them to use their native languages and asking others to translate, slowing down the fast-paced tempo of the classroom, creating an atmosphere in which language variation in the community of discourse is an accepted fact of life -- these methods can apply to bilingual learners. After all, we should not need to reinvent the wheel for every population.

But also, bilingual educators should develop models for teaching mathematics to bilingual students that are not derivative. Lisa Delpit (1986) wrote about the dilemmas of a progressive black educator having to zig-zag between what seems to be today's faddish way of teaching and established ways that work for African American students. She wrote about the search for an authentic way of teaching these children that combines what is successful with them with these emerging developments. As these models are developed, they should inform what occurs in school mathematics. Interestingly, the teachers in Cheche Konnen (Warren & Rosebery, 1990) were not certified in science; they were bilingual teachers who must have used their own knowledge of their students to help guide and develop their program. Now that program is being exported, from bilingual classrooms to the entire school.

Assessment and Evaluation

The issues raised earlier vis-à-vis authentic assessment become increasingly complex when they relate to the bilingual learner. There are, of course, some simple techniques in bilingual education for enhancing student understanding of a task. These include rewriting and simplifying language, using familiar contexts, and providing concrete referents. Difficulties will become immediately obvious with the development and application of scoring rubrics and of performance standards.

On one hand, rubrics that are based on learning theories will have to be modified to ensure that evidence concerning actual knowledge of mathematics is obtained and that evidence is not confounded with difficulties that some children may have expressing themselves in English. On the other hand, if unified theories for learning mathematics and a second language could be developed, it might be possible to create tasks and rubrics based on those theories.

Bilingual educators have had much experience in using scoring rubrics that rely on judgments about the quality of linguistic performance, viz., the assessment of oral language proficiency. The Language Assessment Scales (De Avila & Duncan, 1981; Duncan and De Avila, 1986, 1987) include the collection of speech samples, as does the Functional Language Assessment for older students (Hamayan, Kwiat, & Perlman, 1985). Scoring of these samples is against English-speaking norms, which would be the equivalent of calibrating performance on mathematics assessment against expert performance.

It may be possible to create unified assessments that serve multiple purposes. For example, someone might read some mathematics problems to an LEP student and ask the student to repeat each problem before solving it. Student repetitions could serve as speech samples that would be scored along lines of proficiency. Theories of short-term memory for bilingual populations might provide a means for scoring the same sample along lines of what the student understood about the problem. Then, the problem's solution could be scored as an indicator of the student's actual mathematical knowledge. An additional value to such an approach -- besides its cost effectiveness -- is that language proficiency would be assessed using language similar to what the student would encounter in the classroom.

There is reason for concern about the new rules for assessment and culturally diverse populations. There is an increasing body of evidence that children are socialized according to diverse norms when it comes to how performance on socially desirable tasks is evaluated (Deyhle, 1987; Fillmore, 1989, 1990). Deyhle (1987) documented how American Indian children are socialized to judge for themselves when a task has been learned well enough to be put on display and that judgments about performance quality are highly inappropriate. Hence, assessment tasks that ask such students to show all of their work or timed tasks may be met with resistance by some minority-language students.

The NCTM (1989) recommendations for program evaluation are well taken. Outcome data are not adequate for evaluating the quality of the mathematics programs that students encounter. This recommendation takes on particular importance in view of the traditional reluctance for bilingual-education-program evaluation and research to look at the quality of the school mathematics that LEP students encounter. Mathematics educators will need to understand, however, that bilingual-education-program evaluation needs to consider not just the academic aspects of a program but also language development.

Such evaluation efforts would be helped were there to be some clearly articulated theories that look for points where programs can develop both mathematics and English language proficiency (and also the native language, as appropriate), places where one aspect should take precedence, and places where there must be trade offs.

Concluding Comments

Program evaluation is, in part, an issue of asking about effectiveness. One could liken it to asking about a car's gas mileage to see whether it is worth buying. If so, then the evaluation of mathematics programs for bilingual learners in a time of reform is akin to asking not only about gas mileage but also asking for the answer while the car is running and simultaneously being rebuilt from the ground up -- not an easy task.

There is much worth in the current school mathematics reform movement. That assumption is tacit insofar as I refer to the moving target and am arguing that bilingual education programs need to begin to shift their own goals in light of the new goals for mathematics. Also, there is much of worth in previous bilingual-education-program evaluation. I argue that the conversation needs to go both ways; that people in the education of LEP students should adapt but also should be unafraid of developing ways for teaching the bilingual learner that are not derivative; and that in the history of bilingual education research there have been debates that are similar to those currently found in mathematics education.

We should not think that all debates have been resolved or that most of the technical questions have been answered. Indeed, those efforts are merely beginning. And insofar as there remain open issues and questions, there is room for those who are involved in the education of LEP students to affect that movement through our own practice and research.

Notes

[1] Or, if one follows Baker and de Kanter's (1983) criteria, Does the program work better than any other alternative program?

[2] This report has been very criticized for its many technical flaws (Secada, 1990a; Willig, 1985).

[3] These models are structured-English-immersion strategy, early-exit and late-exit transitional bilingual education programs. They are defined and operationalized in Ramirez (1986) and in Ramirez, Yuen, Ramey and Pasta (1991).

[4] English monolingual students, English-only Hispanics, Spanish-only Hispanics, English-Spanish bilingual Hispanics, Italian-English bilinguals, French-English bilinguals, and German-English bilinguals.

[5] Given Carpenter's (1985) arguments about the knowledge that children enter school with and the results of the program known as Cognitively Guided Instruction (Carpenter, Fennema, Peterson, Chiang, & Loef, 1990; Peterson, Fennema, & Carpenter, in press), maybe this goal should be changed to each student should REMAIN confident in her or his abilities to do mathematics.

[6] In her comments on an earlier draft of this paper, Mary Lindquist raises an additional point. Even the most authentic tasks may suffer from a problem with "so what." For all of our efforts to design such tasks, students (or adults for that matter) may still reject them as uninteresting or as irrelevant. In her comments, for example, Lindquist pointed out that she moved from Madison to someplace else, but she did not engage in sorts of mathematical work that I have proposed as an authentic task elsewhere in this manuscript.

[7] Again, I would like to acknowledge Mary Lindquist's comments on this point. As she notes, part of the power of mathematics comes from our assuming that things are -- for all practical purposes -- like these idealized shapes. We solve the problem in the ideal setting and then apply it to the real world. While granting the need to assume an ideal world -- but only sometimes -- my other objections stand. Who would let a Ben and Jerry's ice cream cone melt all the way? And wouldn't the cone leak anyway?

[8] This graphical representation of real data also has been reported by Warren and Rosebery in Cheche Konnen (1990a, 1990b).

[9] This is not to argue that tasks like these should not serve instructional purposes. Indeed, problems like this one might make college a more viable after high school option for students who seldom, if ever, think of it as an option. While a worthy instructional task, this task is too biased as a stand-alone assessment task to be useful.

[10] I would like to acknowledge Sherian Foster and Matthew Weinstein's contributions to those efforts.

[11] Recall that her teacher had told her to divide by 30.

[12] I would not be so distressed were half of Lampert's class LEP. Then, one could argue that the techniques for creating discourse communities are being invented and refined, and that they do not result in a disproportionate exclusion of students. Lampert does not write about this.

References

American Association for the Advancement of Science (1989). Science for all Americans. Washington, DC: Author.

American Mathematical Society (1990). Mathematicians and education reform (CBMS issues in mathematics education, volume 1). Providence, RI: Author.

Andersson, T. & Boyer, M. (1978). Bilingual schooling in the United States (2nd ed.). Austin, TX: National Educational Laboratory Publishers, Inc.

Archbald, D. A. & Newmann, F. M. (1988). Beyond standardized testing: Assessing authentic academic achievement in the secondary school. Reston, VA: National Association of Secondary School Principals.

Baker, K. A., & de Kanter, A. A. (1983). Federal policy and the effectiveness of bilingual education. In K. A. Baker & A. A. de Kanter (Eds.), Bilingual education: A reappraisal of federal policy (pp. 33-86). Lexington, MA: Lexington Books.

Ball, D. L. (1990). Halves, pieces, and twoths: Constructing and using representational contexts in teaching fractions. East Lansing, MI: National Center for Research on Teacher Education, School of Education, Michigan State University.

Boruch, R. F. & Cordray, D. S. (1980). An appraisal of educational program evaluations: Federal, states, and local agencies (Report to U.S. Department of Education, contract no. 300-79-0467). Evanston, IL: Northwestern University.

Bratt-Paulston, C. (1980). Bilingual education: Theories and issues. Rowley, MA: Newbury House.

Brophy, J. E. (1991). Effective schooling for disadvantaged students. In M. S. Knapp & P. M. Shields (Eds.), Better schooling for children of poverty: Alternatives to conventional wisdom (pp. 211-234). Berkeley, CA: McCutchan.

Brownell, W. A. (1928). The development of children's number ideas in the primary grades (Supplementary educational monograph no. 35). Chicago: University of Chicago Press.

Campbell, D. T. & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Cantoni-Harvey, G. (1987). Content-area language instruction: Approaches and strategies. Reading, MA: Addison-Wesley.

Carpenter, T. P. (1985). Learning to add and subtract: An exercise in problem solving. In E. A. Silver (Ed.), Teaching and learning mathematical problem solving: Multiple research perspectives (pp. 17-40). Hillsdale, NJ: Lawrence Erlbaum.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. & Loef, M. (1990). Using knowledge of children's mathematical thinking in classroom teaching: An experimental study. American Educational Research Journal, 26(4), 499-531.

Carpenter, T. P. & Fennema, E. (in press). Cognitively guided instruction: building on the knowledge of students and teachers. International Journal of Educational Research (Special issue on The reform of school mathematics in the United States, edited by W. G. Secada).

Carpenter, T. P. & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. Journal for Research in Mathematics Education, 15(3), 179-202.

Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), Handbook of research on teaching (2nd ed., pp. 432-463). New York: Macmillan.

Chipman, S. f. & Thomas, V. G. (1987). The participation of women and minorities in mathematical, scientific, and technical fields. In E. Z. Rothkopf (Ed.), Review of Research in Education (Vol. 14, Chapter 9, pp. 387-430). Washington, DC: American Educational Research Association.

Chamot, A. U. & O'Malley, M. (1987). Mathematics book A: Learning strategies for problem solving (Language development through content series). Reading, MA: Addison Wesley.

Clark, C. M. & Peterson, P. L. (1986). Teachers' though processes. In M. C. Wittrock (Ed.), Handbook of research on teaching (2nd ed., pp. 255-296). New York: Macmillan.

Collins, A. (1991). Cognitive apprenticeship and instructional technology. In L. Idol & B. F. Jones (Eds.), Educational values and cognitive instruction: Implications for reform (pp. 121-138). Hillsdale, NJ: Lawrence Erlbaum.

Collins, A. Hawkins, J., & Carver, S. M. (1991). A cognitive apprenticeship for disadvantaged students. In B. Means, C. Chelemer, & M. S. Knapp (Eds.), Teaching advanced skills to at-risk students: Views from research and practice (pp. 216-243). San Francisco, CA: Jossey Bass.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally.

Cook, T. D., & Shadish, W. R., Jr. (1986). Program evaluation: The worldly science. Annual Review of Psychology, 37, 193-232.

Crandall, J. A. (Ed.) (1987). ESL through content-area instruction: Mathematics, science, social studies. Englewood Cliffs, NJ: Prentice Hall.

Crandall, J. A., Dale, T. C., Rhodes, N. C., & Spanos, G. (1987). English language skills for basic algebra. Englewood Cliffs, NJ: Prentice Hall.

Crandall, J. A., Dale, T. C., Rhodes, N. C., & Spanos G. (in press). The language of mathematics: The English barrier. Proceedings of the 1985 Delaware Symposium on Language Studies (VII). Newark: University of Delaware Press. [Available from the authors at the Center for Applied Linguistics, Washington, DC.]

Cronbach, L. J. (and Associates) (1980). Towards reform of program evaluation. San Francisco: Jossey Bass.

Dale, T. C. & Cuevas, G. J. (1987). Integrating language and mathematics learning. In J. A. Crandall (Ed.), ESL through content-area instruction: Mathematics, science, social studies (pp. 9-54). Englewood Cliffs, NJ: Prentice Hall.

Danoff, M. N. (1978). Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program: Overview of study and findings. Palo Alto, CA: American Institutes for Research. [Also available through ERIC Document Reproduction Service.]

Danoff, M. N., Coles, G. J., McLaughlin, D. H., & Reynolds, D. J. (1977-1978). Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program (3 vols.). Palo Alto, CA: American Institutes for Research. [Also available through ERIC Document Reproduction System.]

De Avila, E. A., Cohen, E. G., & Intili, J. K. (1982). Improving cognition: A multicultural approach. In S. S. Seidner (Ed.), Issues of language assessment: Foundations and research. Springfield: Illinois State Board of Education.

De Avila, E. A. & Duncan, S. E. (1981). A convergent approach to oral language assessment: Theoretical and technical specification on the Language Assessment Scales (LAS) Form A (Stock 621). San Rafael, CA: Linguametrics.

De Avila, E. A., Duncan, S. E., Navarrete, C. (1987). Finding out/ Descubrimiento. Northvale, NJ: Santillana.

Delpit, L. D., (1986). Skills and other dilemmas of a progressive Black educator. Harvard Education Review, 56(4), 379-385.

Deyhle, D. (1987). Learning failure: Tests as gatekeepers and the culturally different child. In H. Trueba (Ed.), Success or failure? Learning and the language minority child (pp. 85-108). Rowley, MA: Newbury House.

Duncan, S. E. & De Avila, E. A. (1986). Pre-LAS user's manual (Form A). San Rafael, CA: Linguametrics.

Duncan, S. E. & De Avila, E. A. (1987). Pre-LAS Espanol user's manual (Form A). San Rafael, CA: Linguametrics.

Epstein, N. (1977). Language, ethnicity and the schools: Policy alternatives for bilingual-bicultural education. Washington, DC: George Washington University, Institute for Educational Leadership.

Fillmore, L. W. (1989). Language and cultural issues in early education. Manuscript available from the author, University of California, Berkeley.

Fillmore, L. W. (1990). Latino families and the schools. Remarks prepared for the Seminar on California's Changing Face of Race Relations: New Ethics in the 1990's, Sacramento, CA. Available from the author, University of California, Berkeley.

Fillmore, L. W. & Valadez, C. M. (1986). Teaching bilingual learners. In M. C. Wittrock (Ed.), Handbook of research on teaching (2nd ed., pp. 648-685). New York: Macmillan.

Frankenstein, M. (1989). Relearning mathematics: A different third R--radical maths. London: Free Association Books.

Frankenstein, M. (1990). Incorporating race, class and gender issues into a critical math literacy curriculum. Journal of Negro Education, 59(3), 336-347.

Freeman, D. J. & Porter, A. C. (1989). Do textbooks dictate the content of mathematics instruction in elementary schools? American Educational Research Journal, 26, 403-421.

Good, T., Grouws, D., & Ebmeier, H. (1983). Active mathematics teaching. New York: Longman.

Government Accounting Office (1987). Bilingual education: A new look at the research evidence (Briefing report to the Chairman, Committee on Education and Labor, House of Representatives. GAO/PEMD-87-12BR). Washington, DC: Author.

Hakuta, K. (1986). Mirror of language. New York: Basic Books.

Hamayan, E. V., Kwiat, J. A., & Perlman, R. (1985). The identification and assessment of language minority students: A handbook for educators. Arlington Heights: Illinois Resource Center.

Heath, S. B. (1986). Sociocultural contexts of language development. In Bilingual Education Office, California State Department of Education (Eds.), Beyond language: Social and cultural factors in schooling language minority students (pp. 143-186). Los Angeles: Evaluation, Dissemination and Assessment Center, California State University-Los Angeles.

Hiebert, J. (Ed.) (1986). Procedural and conceptual knowledge: The case of mathematics. Hillsdale, NJ: Lawrence Erlbaum.

Hiebert, J. (in press). Reflection and communication: Cognitive considerations in school mathematics reform. International Journal of Educational Research (Special issue on The reform of school mathematics in the United States, edited by W. G. Secada).

Idol, L. & Jones, B. F. (Eds.) (1991). Educational values and cognitive instruction: Implications for reform. Hillsdale, NJ: Lawrence Erlbaum.

Idol, L., Jones, B. F., & Mayer, R. E. (1991). Classroom instruction: The teaching of thinking. In L. Idol & B. F. Jones (Eds.), Educational values and cognitive instruction: Implications for reform (pp. 65-120). Hillsdale, NJ: Lawrence Erlbaum.

Jones, B. F. & Idol, L. (Eds.) (1990). Dimensions of thinking and cognitive instruction. Hillsdale, NJ: Lawrence Erlbaum.

Kaput, J. J. (1991, May 15). Forming a research and development agenda for algebra and quantitative analysis (Discussion draft). Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Educational Research, National Center for Research in Mathematical Sciences Education.

Kennedy, M. M., Birman, B. F., & Demaline, R. E. (1986). The effectiveness of Chapter 1 services. Washington, DC: Government Printing Office.

Kennedy, M. M., Jung, R. K., & Orland, M. E. (1986). Poverty, achievement, and the distribution of compensatory education services. Washington, DC: Government Printing Office.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving arithmetic word problems. Psychology Review, 92, 109-129.

Lajoie, S. (1991). A framework for authentic assessment in mathematics. NCRMSE Research Review: The Teaching and Learning of Mathematics, 1(1), 6-12. [Available from the National Center for Research in Mathematical Sciences Education, Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison]

Lampert, M. (1988). The teacher's role in reinventing meaning of mathematical knowledge in the classroom (Research series no. 186). East Lansing, MI: Institute for Research on Teaching, College of Education, Michigan State University.

Lampert, M. (199a). Connecting inventions with conventions. In L. P. Steffe & T. Wood (Eds.), Transforming children's mathematics education (pp. 253-265). Hillsdale, NJ: Lawrence Erlbaum.

Lampert, M. (1990b). When the problem is not the questions and the solution is not the answer: Mathematical knowing and teaching. American Educational Research Journal, 27(1), 29-64.

Lane, S. (1991). The conceptual framework for the development of a mathematics assessment instrument for QUASAR. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Lave, J. (1988). Cognition in practice. New York: Cambridge University Press.

Lindblom, C. E. & Cohen, D. K. (1979). Usable knowledge: Social science and social problem solving. New Haven, CT: Yale University Press.

Mackey, W., & Beebe, V. N. (1977). Bilingual schools for a bicultural community. Rowley, MA: Newbury House.

Mathematical Association of America (1989). Reshaping college mathematics. Washington, DC: Author.

Mathematical Association of America (1990). Challenges for college mathematic: An agenda for the next decade. Washington, DC: Author.

Mathematical Association of America (1991). A call for change: Recommendations for the mathematical preparation of teachers of mathematics. Washington, DC: Author.

Mathematical Sciences Education Board, National Research Council (1990). Reshaping school mathematics: A philosophy and framework for curriculum. Washington, DC: National Academy Press.

Mathematical Sciences Education Board, National Research Council (1991). Counting on you: Actions supporting mathematics teaching standards. Washington, DC: National Academy Press.

McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., & Cooney, T. J. (1987). The underachieving curriculum: Assessing U.S. school mathematics from an international perspective. Champaign, IL: Stipes.

Mohan, B. A. (1986). Language and content. Reading, MA: Addison Wesley.

Moll, L. C., Velez-Ibanez, C., & Greenberg, J. (1990). Community knowledge and classroom practice: Combining resources for literacy instruction (A technical report from the Innovative Approaches Research Project). Tucson, AZ: College of Education, University of Arizona. [Also available from Development Associates, Arlington, VA]

Myers, D. E. & Milne, A. M. (1982, October). Mathematics opportunity and exposure: Language minority status and gender (Draft. Prepared for Part C: Math and Language Minority Students Project, U.S. ED contract no. 300-80-0778). Washington, DC: Decision Resources.

National Council of Teachers of Mathematics (1979). Agenda for action. Reston, VA: Author.

National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics (1991). Professional standards for teaching mathematics. Reston, VA: Author.

National Research Council (1989). Everybody counts: A report to the nation on the future of mathematics education. Washington, DC: National Academy Press.

National Research Council (1991). Moving beyond myths: Revitalizing undergraduate mathematics. Washington, DC: National Academy Press.

Oakes, J. (1990a). Opportunities, achievement, and choice: Women and minority students in science and mathematics. In C. B. Cazden (Ed.), Review of Research in Education, 16, 153-222.

Oakes, J. (1990b). Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science. Santa Monica, CA: RAND.

O'Malley, J. M., & Chamot, A. U. (1990). Learning strategies in second language acquisition. New York: Cambridge University Press.

Orum, L. S. (1983). The question of effectiveness: A blueprint for examining the effects of the federal bilingual education program. NABE News, 6, 3-13.

Orr, E. W. (1987). Twice as less. New York: Basic.

Peterson, P. L., Carpenter, T. P., & Fennema, E. (1991). Using children's mathematical knowledge. In B. Means, C. Chelemer, & M. S. Knapp (Eds.), Teaching advanced skills to at-risk students: Views from research and practice (pp. 68-101). San Francisco: Jossey Bass.

Porter, A. (1989). A curriculum out of balance: the case of elementary school mathematics. Educational Researcher, 18(5), 9-15.

Porter, A., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1988). Content determinants in elementary school mathematics. In D. A. Grouws, T. J. Cooney, & D. Jones (Eds.), Effective mathematics teaching (pp. 96-113). Hillsdale, NJ: Lawrence Erlbaum.

Ramirez, J. D. (1986). Comparing structured English immersion and bilingual education: First-year results of a national study. American Journal of Education, 95, 122-149.

Ramirez, J. D., Pasta, D. J., Yuen, S. D., Billings, D. K., & Ramey, D. R. (1991). Final report: Longitudinal study of structured English immersion strategy, early-exit and later-exit transitional bilingual education for language-minority children. Volume II (Contract no. 300-87-0156). San Mateo, CA: Aguirre International.

Ramirez, J. D., Yuen, S. D., & Ramey, D. R. (1991). Final report: Longitudinal study of structured English immersion strategy, early-exit and later-exit transitional bilingual education for language-minority children. Executive summary (Contract no. 300-87-0156). San Mateo, CA: Aguirre International.

Ramirez, J. D., Yuen, S. D., Ramey, D. R., & Pasta, D. J. (1991). Final report: Longitudinal study of structured English immersion strategy, early-exit and later-exit transitional bilingual education for language-minority children. Volume I (Contract no. 300-87-0156). San Mateo, CA: Aguirre International.

Reauthorization of expiring federal elementary and secondary education programs: Bilingual education (Vol. 4. Hearing before the subcommittee on elementary, secondary, and vocational education of the Committee on Education and Labor, House of Representatives. Hearing held in Washington, DC, March 24, 1987. Serial no. 100-5). Washington, DC: Government Printing Office.

Resnick, L. B. (1987a). Education and learning to think. Washington, DC: National Academy Press.

Resnick, L. B. (1987b). Learning in school and out. Educational Researcher, 16(9), 13-20.

Resnick, L. B., & Resnick, D. P. (1991). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement, and instruction (pp. 37-75). Boston: Kluwer.

Romberg, T. A. & Carpenter, T. P. (1986). Research on teaching and learning mathematics: Two disciplines of scientific inquiry. In M. C. Wittrock (Ed.), Handbook for research on teaching (2nd ed., pp. 850-873). New York: Macmillan.

Romberg, T. A., & Stewart, D. (Eds.) (1987). The monitoring of school mathematics (3 vols.). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research, School of Education.

Romberg, T. A., & Tufte, R. (1987). Mathematics curriculum engineering: Some suggestions from cognitive science. In T. A. Romberg & D. Stewart D. (Eds.), The monitoring of school mathematics (Vol. no. 2, Chapter 15, pp. 71-108). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research, School of Education.

Romberg, T. A., Zarinnia, A., & Williams, S. (1989). The influence of mandated testing on mathematics instruction: Grade 8 teachers' perception. Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research, National Center for Research in Mathematical Sciences Education.

Secada, W. G. (1990a). Research, politics, and bilingual education. Annals of the American Association on Political and Social Science, 508, 81-106 (Special issue on Bilingual Education edited by C. Cazden and C. Snow).

Secada, W. G. (1990b). The challenges of a changing world for mathematics education. In T. J. Cooney (Ed.), Teaching and learning mathematics for the 1990s. (1990 Yearbook of the National Council of Teachers of Mathematics pp. 135-143). Reston, VA: NCTM.

Secada, W. G. (1991a). Degree of bilingualism and arithmetic problem solving in Hispanic first graders. Elementary School Journal, 92(2), 211-229.

Secada, W. G. (1991b). Agenda setting, enlightened self-interest, and equity in mathematics education. Peabody Journal of Education, 66(2), 22-56. [Special issue on Needed: An agenda for equity in mathematics education, edited by W. G. Secada & M. R. Meyer]

Secada, W. G. (in press). Race, ethnicity, social class, language, and achievement in mathematics. In D. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 623-660). New York: Macmillan.

Secada, W. G. & Meyer, M. R. (Eds.) (1991). Needed: An agenda for equity in mathematics education [Special issue]. Peabody Journal of Education, 66.

Siegler, R. (1991). Children's learning (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Silver, E. A. (in press). Assessment and mathematics education reform in the United States. International Journal of Educational Research (Special issue on The reform of school mathematics in the United States, edited by W. G. Secada).

Simich-Dudgeon, C., McCreedy, L., Schleppegrell, M. (1988/89). Helping limited English proficient children communicate in the classroom. A handbook for teachers (Program Information Guide Series no. 9). National Clearinghouse for Bilingual Education.

Skutnabb-Kangas, T. (1988). Multilingualism and the education of minority children. In T. Skutnabb-Kangas & J. Cummins (Eds.), Minority education: From shame to struggle (pp. 1-44). Clevedon, England: Multilingual Matters.

Spanos, G., Rhodes, N. C., Dale, T. C., & Crandall, J. (1988). Linguistic features of mathematical problem solving: Insights and applications. In R. R. Cocking & J. P. Mestre (Eds.), Linguistic and cultural influences on learning mathematics (pp. 221-240). Hillsdale, NJ: Lawrence Erlbaum.

Stallings, J. A. & Stipek, D. (1986). Research on early childhood and elementary school teaching programs. In. M. C. Wittrock (Ed.), Handbook of research on teaching (2nd ed., pp. 727-753). New York: Macmillan.

Stanic, G. M. A., (1991). Social inequality, cultural discontinuity, and equity in mathematics. Peabody Journal of Education, 66(2), 57-71. [Special issue on Needed: An agenda for equity in mathematics education, edited by W. G. Secada & M. R. Meyer]

Steen, L. A. (Ed.) (1990). On the shoulders of giants: New approaches to numeracy. Washington, DC: National Academy Press.

Steen, L. A. (1991). Mathematics for all Americans. In T. J. Cooney (Ed.), Teaching and learning mathematics in the 1990s (1990 Yearbook of the National Council of Teachers of Mathematics, pp. 130-134). Reston, VA: NCTM.

Stein, C. B., Jr. (1986). Sink or swim: The politics of bilingual education. New York: Praeger.

Stenmark, J. K. (1989). Assessment alternatives in mathematics: An overview of assessment techniques that promote learning. Berkeley, CA: University of California, EQUALS, Lawrence Hall of Science.

Stevenson, H. W., Lummis, M., Lee, S-Y., & Stigler, J. W. (1990). Making the grade in mathematics: Elementary school mathematics in the United States, Taiwan, and Japan. Reston, VA: National Council of Teachers of Mathematics.

Stigler, J. W., Lees, S-Y., & Stevenson, H. W. (1990). Mathematical knowledge: Mathematical knowledge of Japanese, Chinese, and American elementary school children. Reston, VA: National Council of Teachers of Mathematics.

Stigler, J. W., & Stevenson, H. W. (1991, Spring). How Asian teachers polish each lesson to perfection. American Educator, 12-20, 43-47.

Tallmadge, G. K., Lam, T. C. M., Gamel, N. N. (1987a). The evaluation of bilingual education programs for language-minority, limited-English-proficient students: A status report with recommendations for future development (Phase I report). Mountain View, CA: RMC Research Corporation.

Tallmadge, G. K., Lam, T. C. M., Gamel, N. N. (1987b). Bilingual education evaluation system: Users' guide. (2 vols.). Mountain View, CA: RMC Research Corporation.

Tikunoff, W. J. (1985). Applying significant bilingual instructional features in the classroom. Washington, DC: National Clearinghouse for Bilingual Education.

Warren, B. & Rosebery, A. S. (1990). Cheche Konnen: Collaborative scientific inquiry in language minority classrooms (2 vols.). Cambridge, MA: Technical Education Research Association. [Also available from Development Associates, Arlington, VA]

White, S. H., & Buka, S. L. (1987). Early education: Programs, traditions, and policies. *Review of Research in Education*, *14*, 43-91.

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, *55*, 269-317.

Zarinnia, E. A., & Romberg, T. A. (1987). A new world view and its impact on school mathematics. In T. A. Romberg & D. A. Stewart (Eds.), *The monitoring of school mathematics* (Vol. 1, pp. 21-61). Madison: Wisconsin Center for Education Research, University of Wisconsin.

Zentella, A. C. (1981). Ta bien: You could answer me in cualquier idioma. Puerto Rican codeswitching in bilingual classrooms. In R. Duran (Ed.), *Latino language and communication behavior* (pp. 109-131). Norwood, NJ: Ablex.

Response to Walter Secada's Presentation

Penelope L. Peterson
University of Michigan

Following up on Walter's comments, I just have to say that all you need to know about me can really be summarized by the fact that I was on the faculty at the University of Wisconsin--Madison for 11 years, and I left. I am now at Michigan State University, quite happily by the way, although my years at Madison were very productive.

At Michigan State University, we have really an outstanding and very interesting group of scholars working on problems of reform of teaching and teacher education in schools. Much of what I am going to say and my ideas and my thinking have been influenced profoundly by my conversations with my colleagues within this community of learners, teachers, and researchers that we have created in the College of Education at Michigan State University. Specifically, I would like to acknowledge the contributions to my own thinking and learning that have resulted from ongoing conversations over the last five years with Deborah Ball, David Cohen, Patrick Dickson, Magdalene Lampert, Sarah McCarthey, Richard Prawat, Ralph Putnam, and Suzanne Wilson.

Our Dean, Judy Lanier, has been influential in creating this thoughtful community of learners, teachers, and scholars in our college. And so I would like to start out with a little metaphor that Judy Lanier has used to talk about this whole problem of assessment and to raise questions about the idea that many people have that "assessment will drive instruction." Judy questions this drive to construct a national test to measure the progress toward reform in education in our nation's schools. Lanier compares our race toward reform with our race to make it to the moon in the 1960s, and she queries: By designing a national test to measure the progress of reform, isn't it a bit like setting the goal to make it to the moon, designing a terrifically big new telescope to see if we made it there, but doing nothing in between?

A major message of Walter Secada's paper is that there is a lot "in between" that needs to be considered seriously. There is a lot in between that we need to think about and take account of if we are really to measure and understand education change and progress. We need to think hard about some of the elements that Secada has pointed out.

I would like to situate my remarks within the context of reform in mathematics education because mathematics education is really the context for Walter's remarks and for Mary Lindquist's comments as well. And, as Walter points out in his paper, we in the mathematics education community are perceived to have a coherent vision for reform. This vision encompasses and extends from the standards of the National Council of Teachers of Mathematics (NCTM) (1989; 1991) to include ongoing reform efforts of members and affiliates of that organization. But it also encompasses the mathematics education reform efforts that are going on in states such as California with the California Mathematics Framework (1987; 1992) that proposes a new and ambitious vision of mathematics instruction. In their remarks, both Walter Secada and NCTM President, Mary Lindquist, have done a nice job of summarizing that vision.

That vision interweaves four important elements. One is that there are new goals for students' learning of mathematics that move beyond computation. The second element is a significant revision in the K-12 mathematics curriculum -- new topics are added, and others are eliminated or reduced. Third, this reform vision really call for a different kind of pedagogy. An important idea is that how mathematics is taught shapes what students learn. Consequently, the reform proposals call for students to talk much more and teachers to talk less, for students to make conjectures and arguments, and for teachers to skillfully direct and moderate students' investigations. Finally, the proposals call for attention to the mathematics learning of all students, African-American, Hispanic, and female students as well as white males.

Now to pick up on that and to quote from Walter's paper: "The shifting of goals and visions for school mathematics has profound implications for the education of LEP students. Assume, for example, that we actually achieved the goals for mathematics that are found at least tacitly in current evaluation practices. Would this be a real success, or would it not be a pyrrhic victory? Were we to succeed in meeting the mathematics goals that are found in current tests, LEP students would become computational wizards, but would be unable to engage in the sorts of mathematical activities that their English-proficient peers would engage in routinely during their own schooling. The target has shifted: the evaluation of school mathematics for LEP students needs to shift as well. Conversely, the mathematics reform movement has failed to pay serious attention to the education of diverse learners....Unfortunately, the new Standards for school mathematics curriculum and its teaching do not include checks to ensure that they will, in fact, apply to everyone, and that resultant practice will meet the diverse needs of this country's LEP students."

I think that one thing that is clear from Walter's paper and from Mary's remarks is that the problems with mathematics instruction are systemic, and that achievement of these ambitious goals will require changes in curriculum, assessment, policies, and structures at all levels of the system from the state to the district to the school to the classroom. What I would like to focus on here is what I see as the invisible actor in Secada's paper, but perhaps the key person in systemic reform -- the teacher. What I have to say is intended to embellish on the arguments that Walter has made in his paper.

In Secada's concluding comments, he uses an apt metaphor to reveal a major difficulty that reformers face. Secada contends that "the evaluation of mathematics programs for bilingual learners in a time of reform is akin to asking not just about a car's gas mileage [to see whether it is worth buying], but asking new questions and asking for the answer while the car is running and simultaneously being rebuilt from the ground up -- not an easy task."

Not an easy task, I agree, but a very apt metaphor. In fact, a similar metaphor that we are fond of using at Michigan State is the idea that, as educators involved in reform at all levels, we are trying to sail a boat while we are building it -- the same idea as driving the car while you're building it from the ground up. But, as I read this metaphor in Secada's paper (and maybe it's because the focus of my research has always been on teachers and teaching and these are what I spend my life thinking about) I just kept thinking -- but it all depends on who is driving the car. What is missing for me in this metaphor is the driver who is driving the car while rebuilding it. The most important driver right now in our American schools and in our nation's classrooms is the teacher. And, what I would like to spend my fifteen minutes talking about is the teacher because I think without teacher support, without active participation on the part of teachers, without profound changes in teacher's beliefs, knowledge, thinking, understanding and expectations, little is going to change. Teachers are the critical mediators of student's mathematics learning, and teachers are the critical agents of this reform.

But teachers are in a difficult position, a very difficult position indeed; and anything I say today is not meant in any way to berate teachers. On the contrary, what I think we need to do is figure out how to help and support teachers. Teachers face incredible challenges. Take, for example, the case of mathematics. Teachers are products of the kinds of classrooms that are currently under fire. The mathematics education reforms invite teachers to construct quite a different kind of teaching and learning, yet they themselves never experienced that kind of mathematics teaching and learning. Further, teachers have not experienced the kind of mathematics that reformers are talking about them teaching. It is unclear whether any of us have ever experienced that. Remember, again, this is a car we're rebuilding as we're driving it along or a boat that we're constructing as we try to sail it.

It is a profound dilemma for the teacher in the classroom, and I would like to propose that it is a profound dilemma for us. I want to spend some of the rest of my time talking about an actual case of a teacher -- an authentic case. I would like to tell you a story about Cathy Swift, a California teacher whom I have been following for three years. I would argue that Cathy Swift is typical of many teachers out there and, because of that, we need to try to understand what she has been going through.

In addition to my personal judgment that Ms. Swift is typical, I have statistical data that placed Cathy Swift among a modal cluster of teachers when we surveyed 493 elementary teachers in California, Florida, and Michigan about their current goals and activities in teaching mathematics. (See Peterson, Putnam, Vredevoogd, and Reineke, in press). Cluster analysis of teachers' survey responses yielded five clusters of teachers: (a) primary teachers who had students use manipulatives extensively; (b) *Math their Way* teachers who had students use manipulatives and discuss problem solving extensively; (c) modal teachers whose profile reflected a softened version of drill-and-practice teachers; (d) drill-and-practice teachers; and (e) teachers in the expert cluster whose profile represented a balanced version of the *Math their Way* teachers' profile. Cathy Swift's survey response fell into the modal cluster of teachers. After we conducted this survey, we began to do case studies of twenty-four elementary teachers in the state of California in which we went into their classrooms and interviewed the teachers and observed their mathematics teaching practice.

These case studies are part of a longitudinal study of policy and practice that I have been conducting with several Michigan State colleagues in which we have been examining the relationship between the state level reform in mathematics in California and classroom practice. Building on the notion of systemic reform, the California mathematics education reform has several elements. One element is the California Mathematics Framework (California State Department, 1985; 1992) which lays out the new vision of mathematics, learning, and teaching aimed at "teaching mathematics for understanding." The second element is the selection of textbooks or the design of curriculum materials aligned with the Framework. A third element is the construction of new assessments of students' mathematics learning that are aligned with the Framework and the texts. In our study, we are interested in what teachers are doing when one looks behind the classroom door. Our picture of what we found in teachers' classrooms came out in the Fall, 1990, issue of Educational Evaluation and Policy Analysis (EEPA) in which we provided case studies of five different elementary teachers' classrooms in three different California school districts. (See Ball, 1990; Cohen, 1990; Peterson, 1990; Wiemers, 1990; and Wilson, 1990).

The teacher that I wrote about in my EEPA case study is a teacher whom I call Cathy Swift. Cathy Swift is teaching in a school district that has 118,000 students. It is a very large urban district. The large urban elementary school in which Ms. Swift teaches has an extensive minority population; many immigrants come into this school; many of the students are Limited English Proficient; and most of the students (90 percent) qualify for free or reduced lunch. Substantial ethnic and linguistic diversity exists within the school with 20 different languages being spoken by children who are enrolled. Signs posted in the building and information for families in the staff lounge are in English, Spanish, Lao, Vietnamese, Cambodian, and Hmong. In my initial case study, I summarized my impressions of what I saw as "a smoothly and swiftly-paced model lesson in the tradition of effective teaching for basic skills -- warm-up, review, and seatwork, with continuous monitoring by the teacher, direct instruction, and directed prompting when a student needs help." In other words, I saw Cathy enact marvelous direct instruction lessons in the tradition of active mathematics teaching.

That was in the 1988-89 school year. In my case analysis, I argued that one reason that Cathy taught the way she did was because she was teaching within a model that the school district had adopted called the Achievement for Basic Skills (ABS) model. This model was based on master learning ideas where teachers were given pacing charts, mastery tests to assess students, and additional worksheets to use for remediation when students failed to pass the mastery tests. Teachers were told to use direct instruction, and they had to turn in their pacing charts and their scores on their mastery tests to a mentor teacher in their school who reported them to the ABS office in the district. I argued that Ms. Swift's practice was framed by having to teach within that context. Now I would like to tell you what I saw in Cathy Swift's classroom the following year when I went.

When I returned to Ms. Swift's classroom, it was the 1989-90 school year, and Swift had elected to switch to teaching a group of students that were limited English proficient. She had a class called a "sheltered" class. Although none of her students had English as their native language, Cathy was supposed to teach the class in English, and she did. She taught in a small bungalow that had been added to the school because the school was overcrowded, having been built for 300 students and now housing more than 900 students. When I entered the bungalow, I was struck by Ms. Swift's class -- thirty-one faces looked up at me that varied in shades from yellow to brown to black. The three white faces in the room were Cathy and I and the thirty-second student who was a fair-skinned white girl with bright red hair who was a native Russian speaker. Although Cathy herself speaks no languages other than English, she told me that she had decided to teach this fourth grade sheltered class because she wanted to get out of the "restrictiveness" of the ABS model, and teachers of sheltered classes were not required to follow the ABS model. Immediately, I thought to myself: "Good! Great! We're going to see interesting mathematics teaching now, right? Fantastic kinds of things."

Cathy began by telling me that what she thought LEP students need is lots of "hands-on" experiences with mathematics, and they need a lot of "active involvement." As I watched Cathy Swift teach a lesson to her LEP students, I saw her attempt to put her ideas into practice. She began with a short review that dealt with "fact families," and then, for the second part of the lesson, she read a book to her students, How much is a million? Ms. Swift read the book aloud and asked her students factual questions that dealt with information in the text. But what was striking was the missed opportunity for asking the students some very interesting questions, such as asking the students to speculate about the size of a million or querying them about what they thought a person might buy with a million dollars.

The last part of Ms. Swift's lesson was, in her words, "a review of place value." Now pretend you were in this classroom situation, and you were sitting there trying to make sense of what was going on, and I will describe to you what was happening was the following. Ms. Swift passed out different colored cards to her fourth-graders. Each card had a number from 0 to 9 written on it, and each child got two cards. The color of each card matched one of the colors of the "places" on the board: the ones' place on the board was beige, the tens' place was pink, the hundreds' place was red, and the thousands' place was blue.

Ms. Swift began the activity by announcing: "I'm going to write a number on the board, and you look at your card. If you have the card that goes in that place, I want you to get up and stand in that place." To demonstrate what she meant, Ms. Swift wrote the number "100" on the board. She wrote a "1" above the red hundreds' place on the board, a "0" above the pink tens' place, and a "0" above the beige ones' place on the board. Then she called on the person with "one hundreds" to come up. Hector announced that he had it so he marched to the board and stood under the red hundreds' place holding his card in front of him. Hector was holding up a dark blue card with a one on it.

Ms. Swift said to Hector, "No, you have the thousands, not the hundreds." Holding up Hector's card, she asked the class, "Does this go in the hundreds place?"

The class chorused in unison, "No!"

Ms. Swift said, "Then, well, who has the hundreds' place?"

One child called out, “the red.”

The child with the “1” on a red card came up and stood beneath the red one hundreds’ place on the board.

Ms. Swift then asked, “Now, what do we have in our tenth place?”

The class chorused in unison, “zero!”

The teacher queried, “Who has that one?” and a child with a pink card with a “0” on it came up and stood beneath the tens’ place at the board.

Finally, Ms. Swift asked, “Okay, who has the ones’ place?”

A girl with a beige card with a “0” on it went to the board and stood under the ones’ place.

Looking at all three children holding their cards at the board, Ms. Swift summarized, “Okay, reds are hundreds, pinks are tens, and beige is the ones’ place. Who can read our number for us?” She called on Belinda who responded correctly, “one hundred.”

Ms. Swift continued the place value activity for several minutes by having the students enact each of several more numbers. As with the above example, the students were “actively involved” in this “hands-on” activity” as the students with the appropriate cards came to the board to represent the places in the number.

Let me summarize what I see as significant in this case of Cathy Swift -- a teacher who is trying very hard to teach mathematics for understanding to her limited English proficient students. Cathy Swift is a thoughtful, hard-working teacher, and a sensitive, compassionate, caring person. She chose to teach in this large, urban overcrowded school with children from a diversity of ethnic, linguistic, and socioeconomic backgrounds; she could have chosen to teach in a less challenging situation. Looking at Swift’s teaching from one perspective of where she was the previous year, she has made significant changes. She has moved beyond the direct instruction model and is engaging in activities that are very much consonant with the mathematics education reform. We saw Cathy attempt to integrate literature into her mathematics teaching by reading a story about numbers to her LEP students. Her students appeared engaged throughout the reading. Further, Cathy is using what she sees as “active involvement” and “hands-on manipulatives” in her mathematics teaching. Cathy thinks of her LEP students as achieving concrete understanding of place value through the kinesthetics of pairing the placement of their body on the color of their card with where they place the number. Yet from the perspective of most mathematics educators, Cathy's understanding and her practice reflect a rather rote conception of place value.

Writers of the California Mathematics Framework and Model Curriculum Guide (California State Department, 1987) would argue that Ms. Swift has really missed the "essential understanding" of place value which they articulate as follows:

"Any number can be described in terms of how many of each group there are in a series of groups. Each group in the series is a fixed multiple(the base of the place value system) of the next smaller group. The place value system requires the act of counting groups as though they were single items. It is this organizational structure that gives us the power to deal with large numbers and small numbers in reasonable ways. Rather than endless, unfathomable series of numbers, we need only the digits zero to nine. By grouping we can think of a hundred as a unit or a trillion as a unit; by subdividing we can think either of one thousandth or one millionth of a unit. We can record very large and very small numbers by using the position of the digit to indicate the group we are using as a unit" (California State Department of Education, 1987, p. 19).

Why did Cathy Swift teach place value the way she did to her fourth-grade class of LEP students? One way of thinking about Swift's practice is that when she ceased to work within the direct instruction model, she was freed from constraints, but she was also left to recreate her classroom practice from her own knowledge, beliefs, and understandings. So what did Cathy Swift do? She attempted to bootstrap up from her knowledge and understandings which she herself admits are incomplete in the area of mathematics. For example, when I asked Cathy about her mathematics course at the liberal arts college she attended, she said, "it was a joke." Cathy Swift acknowledges that she does not know how to teach children to solve problems. Yet like all of us, what Cathy sees and understands is framed within and limited by her own understandings and perspectives so that she sees only what she can see from her own point of view. So when I asked her about the California Mathematics Framework, Cathy said that she had attended a seminar where they "read the framework from cover to cover." "Great! I thought to myself," so I asked out loud, "What did you think about it? Did you have any new insights?" Swift replied, "Well, actually it's a pretty boring, dull document. I guess it just reaffirms what I'm already doing."

Why do I tell this story of Cathy Swift? I tell it to illustrate for you the average teacher's dilemmas within the contexts of this current education reform. Although I have used this one case, I do believe that, in several important ways, Cathy Swift represents the typical elementary teacher. In this case, Swift has moved to teaching LEP students so she faces even greater challenges than the typical teacher of white, middle-class students. Cathy Swift's dilemmas are these: she is being asked to teach a new mathematics that is different from the mathematics she learned; with a new pedagogy that is different from the way she was taught; to achieve new goals different from basic skills; to a new group of students, more diverse than those with whom she attended school and who have certainly more diverse ethnic, linguistic, and social knowledge, backgrounds and experiences than Cathy's own reflect. Cathy Swift is being asked to do all this without being supported and helped to attain the kinds of new knowledge and skills that she will need to do it.

I would argue that these are dilemmas that we cannot just let the Cathy Swifts of the world confront alone. We must confront them as well. As teacher educators, policy makers, administrators and researchers, we must somehow confront these dilemmas with Cathy. If we do not confront these dilemmas and help and support teachers in developing the new knowledge, skills, understanding, and dispositions that they will need to reconstruct the car or build the boat, then we will not need to spend millions of dollars to do a meaningful evaluation of mathematics education of limited English proficient students of the kind that Walter Secada so eloquently described in his paper. We can just reread the research reports of the evaluations that have been done over the last decade. We will not need to do a million-dollar evaluation because if we do not join in confronting teachers' dilemmas with them, then nothing significant will change in the mathematics education of the average American student let alone in the education of the average limited English proficient student.

References

Ball, D. L. (1990). Reflections and deflections of policy: The case of Carol Turner. Educational Evaluation and Policy Analysis, 12(3), 247-259.

California State Department of Education (1985). Mathematics framework for California public schools: Kindergarten through grade twelve. Sacramento, CA: Author.

California State Department of Education (1987). Mathematics Model Curriculum Guide. Sacramento, CA: Author.

California State Department of Education (1992). Mathematics framework for California public schools: Kindergarten through grade twelve. Sacramento, CA: Author.

Cohen, D. K. (1991). A revolution in one classroom: The case of Mrs. Oublier. Educational Evaluation and Policy Analysis, 12, 311-30.

National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics (1991). Professional standards for teaching mathematics. Reston, VA: Author.

Peterson, P.L. (1990). Cathy Swift: Doing more in the same amount of time. Educational Evaluation and Policy Analysis, 12(3), 261-280.

Peterson, P.L., Putnam, R.T., Vredevoogd, J, & Reineke, J. (in press). Profiles of practice: Elementary school teachers' views of their mathematics teaching. In W. G. Secada (Ed.), Researching educational reform: The case of school mathematics in the United States. (Special issue of the International Journal of Educational Research).

Wiemers, N. J. (1990). Transformation and accommodation: A case study of Joe Scott. Educational Evaluation and Policy Analysis, 12(3), 281-292.

Wilson, S. (1990). A conflict of interests: The case of Mark Black. Educational Evaluation and Policy Analysis, 12(3), 293-310.

Response to Walter Secada's Presentation

Mary Lindquist
Columbus College, Georgia

As the President-elect of the National Council of Teachers of Mathematics (NCTM), let me first thank you for the opportunity to attend this conference and to learn from you. In the few minutes that I have this afternoon, I would like to assist Walter Secada--not that he needs much assistance--in presenting the mathematics community's view toward reform in mathematics, to raise a few concerns related to his paper, and to reinforce Walter's discussion of needed steps in collaboration between you and those of us in mathematics education.

First let me say I could not agree more with one premise of Walter's paper, that "evaluation of school mathematics for LEP students needs to shift," and with one of his warnings, and I quote: "if we continue to do more of the same, even if we try to do a better job of it, we may achieve our goals, but they are outdated and inadequate for purposes of preparing LEP students to participate in the world in which they will live their adult lives." In these quotes, all students could be substituted for LEP students...it is not just your problem...it's a problem for all our students. This is why NCTM responded and produced two documents, the Curriculum and Evaluation Standards for School Mathematics and the Professional Standards for Teaching Mathematics.

Let me share some of my views of the vision of these two documents even though Walter has done an excellent job of explaining the view of the mathematics community. I was afraid when he said it was his critical day -- once in a while he gets real critical on lots of issues -- but he was very gentle today. To consider the mathematics community position, return with me to Thomas Popketwitz's talk this morning when he compared the field of change to a baseball field. I have often felt that we in mathematics education have been an outfield; I hope we have made it to shortstop now. Hopefully, this vision of ours is not just a field of dreams. But if there is a field of dreams, you will come and we together can make a difference.

Walter Secada stated the five goals of the NCTM Standards. Let me reiterate them quickly. Students should become mathematical problem solvers, they need to learn to reason mathematically, and to communicate mathematics. The other two goals address the value of mathematics. Do our students value mathematics? Do the students you work with value mathematics? One result from NAEP: eighth graders across the nation as a whole think mathematics is extremely important. When asked, important for whom, however, individual students respond: it's important for somebody else, not for me. If we can achieve this last goal, each child should become confident in his or her ability to do mathematics, we will make progress.

Let me say a few things about equity because Walter does address this issue in the paper and make a confession. Early in my teaching career, I thought I had made it when I got five boys, yes, boys, in calculus. For a long time, many of us in mathematics thought of math as a filter, an exclusive club for only a few. One of the changes today is a relook at that attitude. We have made progress, but we have a way to go. Mathematics is still a filter.

As I have looked at National Assessment of Educational Progress (NAEP) data over the years, it has concerned me greatly that we have not provided the opportunity for everybody to experience a broad curriculum. We have closed the gap of performance on numbers and operations among different groups of students. But the gap still exists in measurement, geometry, and problem solving. It's not because our students can't learn; it's because a lot of them are not given the opportunity to learn.

The standards consider this broader view of mathematics. That's one reason we have an algebra standard in K-4. It opens the door to everybody rather than make algebra a cut-off. We need to do more than teach mathematics, year after year, that can be done with a $3.95 calculator. We need more math, and we need different mathematics. I think one of the most exciting aspects of the new vision is the emphasis on communication. At each group of grade levels (K-4, 5-8, and 9-12), there is a standard on communication in mathematics. These are standards that each of you may want to read because they do tie our two interests together.

I want you all to think for a minute of the computation exercise (you might want to write it down) 5 3/4 divided by half. What do our students do with that? There are many of our students that give us an answer. But does it make sense? Can they give you a situation -- I don't know if this is authentic or not -- but can they even give a situation that includes any language other than five and three over four divided by half. What meaning does it have? When they get an answer, does it make sense?

I know the first thing many students say is "five and three quarters, I have to change that to an improper fraction and now what do I do? I think I do something in a circle with those fractions. Divided by half, I think I flip something." There's no meaning there. There's no language there that gives meaning. I believe that almost all of our students across the nation could solve this problem if it were set in a context. Think about it yourself. If you had, and I know that I won't pick the right context, five and three-fourths pies. If you gave a half a pie to each family, to how many families could you give? Well, think about one pie. If you were going to give half to each family, you'd give to how many families? Two. How about two pies? Four. You all are bright, try five. Ten. Gosh, I'm dividing but I get an answer larger than I started with. That doesn't fit the conception of division held by a lot of our children.

Also, you begin to realize that what you did was multiply by two. Maybe there is something to that flipping or inverting. We need to work with the language, and we need to begin with the children's language and build the mathematic language from their language. In summary, my vision of the standards include mathematics that makes sense to all children.

In Walter's paper he talked about a discipline-based task versus an authentic task. I don't think there is a need to be polar. I think we need both. Mathematics is a discipline; we can't leave the mathematics out of our assessment.

I want to examine two examples that he gives in his paper. One example was about moving from Madison. I moved from Madison once, and I never went through all that. How authentic is that problem for 10th graders. I agree wholeheartedly that we have often taught math so students would do better in the next grade. That's ridiculous. We need to have real life, whatever that is. But remember that real life for young children is often fantasy, and for older children it's not our real life. We do need to make mathematics useful or authentic, but we cannot ignore the discipline. I do not mean to return to the 1930s when all mathematics had to be based on use. If you couldn't use it immediately, it was not included. Mathematics is a discipline, and there's some beautiful mathematics that can excite children and help them look at the discipline itself.

I want to argue a little bit with what Walter says about the Connecticut example. Walter says it's not authentic. Let me give you the task: you have an ice cream cone; on top of that cone you put a scoop of ice cream; if the ice cream is a perfect sphere and it all melted down into the cone, would the cone run over? From the students's responses that I have heard, they don't think that's authentic either, but they play along with us, they get engaged in it, and they come up with a variety of ways to solve the problem.

But Walter made one statement that really, really bothered me. He said it was not authentic, because no scoop of ice cream is ever spherical. But that's what we do in math. We make assumptions; that's the basis for the whole discipline of mathematical modeling. We try to simplify the world so that we can work with it. If I assume it's spherical, that's my mathematical assumption to help me work the problem. That assumption doesn't bother me at all. That doesn't make it non-authentic.

What makes it non-authentic to me is, "who cares?" What I would rather solve -- I mean, who ever worried about that problem? Loving ice cream, I would rather know how large a scoop I can get on the top without it falling out. You'll understand these references as you read Walter's paper; it's very readable and very good.

I think we have to be careful not to change one set of problems for another. Here I will sound a little bit defensive because Walter says in his paper that the NCTM Standards really have no authentic examples, no uses -- I may be overstating it a little bit -- no part of the world outside. Yet, as I look through it again, I see many recommendations for collecting data, analyzing data, starting with children's own problems, estimating change, making dog kennels, and so forth. There are efforts to tie mathematics to the outside world and to see its usefulness.

Let me quickly change to assessment since that's been the topic of this conference. I think Walter has put assessment in its proper perspective. There is guidance in the Standards: assessment of students, assessment of teaching, and program evaluation. The main focus is on student assessment that is to improve learning and teaching. The emphasis is on what students can do instead of what they cannot do.

As I work with teachers, some of the most exciting things have happened when they interview their students. At first they're amazed the students can't do and don't understand. What bothers me is that I have not always been able to turn that view around so they can tell me what the students can do. When we do get to that stage, they know what to do next.

I think one issue that Walter raises, whether assessment is the driving force, is crucial. Read the quotes from Ed Silver in Walter's paper about the position that changing assessments will not necessarily change learning and teaching. One of the main issues concerns beliefs and expectations. Until teachers change their beliefs, until society changes its beliefs about mathematics, we will not make progress. You know, it's very acceptable in our nation not to be able to do math. Think about it. Until we change that, I don't know that we'll move forward.

In conclusion, I want to comment on the five steps that Walter recommends taking. I think there are steps that you could take alone, but hopefully we will take together in looking at mathematics. I may be paraphrasing some of these, but I think this is what Walter was saying in his paper.

First of all, set goals for LEP students in mathematics that are in concert with NCTM standards. This does not mean they have to be exactly the same but that you are reaching for the new vision of mathematics.

Second, communicate and continue to do research that will inform the mathematics reform about LEP students. We need to know what you are thinking and what your research is saying. I would add that we also need to work together on the research.

Third, develop samples of contexts that may be unfamiliar to the mainstream culture and ways for teachers to use these. You are the ones that can inform curriculum developers and teachers.

Fourth, help wrestle -- and these were not Walter's words -- help wrestle with assessment issues, especially issues regarding language in cultural context. We need that help in mathematics.

Fifth, encourage program evaluations that focus on the quality of school mathematics that students encounter.

As Walter said, the target's moving. But I think if we work together, we have a much better chance of hitting it than if we work separately.

Science Education as a Sense-Making Practice: Implications for Assessment

Beth Warren and Ann S. Rosebery
Technical Education Research Center, Cambridge

In this paper, we argue for a rethinking of what it means to *do science* in language minority classrooms by putting forward a view of science as a sense-making practice. We then explore the implications of this view for assessment. But before outlining a sense-making perspective on scientific practice, it is helpful to invoke some familiar images of what science is like in many classrooms in order to lay out a few critical connections among teaching, learning, and assessment. Two examples follow, one descriptive of science in many mainstream classes (although the example itself is drawn from a science class outside of the United States) and the other of science in a Chinese bilingual program in California.

> I once witnessed a marvelous science lesson virtually go to ruins. It was a class of young secondary school girls who, for the first time, were let free to handle batteries, bulbs, and wires. They were busy incessantly, and there were cries of surprise and delight. Arguments were settled by "You see?", and problems were solved with, "Let's try!" Hardly a thinkable combination of batteries, bulbs, and wires was left untried. Then, in the midst of the hubbub, the teacher clapped her hands and, chalk poised at the blackboard, announced: "Now, girls, let us summarize what we have learned today. Emmy, what is a battery?" "Joyce, what is the positive terminal?" "Lucy, what is the correct way to close a circuit?"...And Emmy, Joyce and Lucy and the others deflated audibly into silence and submission, obediently copying the diagram and the summary. What they had done seemed of no importance. The questions were in no way related to their work. (Elstgeest, 1985:36-37)

The problem Elstgeest describes is the disjunction between learning and teaching, between what students learn when they engage phenomena directly and what teachers (or curricula or tests) think they should be -- or are -- learning. For a variety of reasons, teaching in many cases is not connected to students' learning, to the sense students make of the world around them. In Elstgeest's example, in fact, the teacher's questions -- which function as a form of evaluation -- act to undermine rather than encourage the students' learning. Focusing narrowly on definitions and correct answers, the questions ignore the students' scientific explorations and efforts to make sense of phenomena.

In language minority classrooms, science -- when it is taught at all -- is often even further reduced (cf. Moll, in press), as illustrated in the following example:

> 9:10 Science T (or prep T) comes in to teach science. She hands out a solar system puzzle and tells students to do it on their own because it is like a quiz. D (NES: Non English Speaking) is playing around. He can't do the handout, so Prep T takes it away. He begins working on his penmanship handout.
>
> The crossword puzzle is too difficult for the NES and LES (limited English speaking) students. I begin working with C (NES) first by explaining the definitions in Chinese (e.g., the largest planet or the ringed planet). He can't recall the word on his own even if he knows the meaning. So I get the encyclopedia volume on the solar system for him to use as a reference book. He is able to answer the first few questions on the crossword puzzle on the planets but gets stuck on the more difficult words. Furthermore, he can't even understand the definition or clue words for the puzzle. (Guthrie, 1985:161-162)

On the surface, at least, this case looks very different from the Elstgeest example and, in some crucial respects, it is. Whereas in the Elstgeest example the students actually got their hands on batteries, bulbs, and wires, in this case the crossword puzzle exercise is abstracted out of any meaningful context of scientific activity. Further, in the second example, science is confounded with English language development. The focus of the exercise is on definition and naming. Students in a class like this memorize the definition of the word "hypothesis" but never experience what it means to formulate or evaluate one.

But, in other respects, Guthrie's example is not so far from Elstgeest's. Underlying the pedagogical approach in both is a view of science as the accumulation of facts, definitions, terminology, and correct procedures. Teachers pose the questions and, more frequently than not, provide the explanations. The Elstgeest example is particularly instructive in this regard because it has at its center hands-on exploration. But hands-on science, it turns out, is not enough. In the absence of a framework for understanding students' scientific sense-making, even the best hands-on curricula can become the occasion for knowledge transmission. It is striking, too, that in both cases teaching doubles as assessment. In the Elstgeest example, the teacher queries the students to see if they have learned the right things: the components of a battery, how to close a circuit. In the Guthrie example, the exercise is set up as a quiz, to assess how much technical English vocabulary the students have acquired.

These images of science are widespread. Recent national and international assessments (Mullis & Jenkins, 1988; McKnight et al.,

761

1987) and calls for reform (AAAS, 1989; Symansky & Kyle, 1990) testify to this fact. More important are the questions raised by these common practices:

- What is the purpose of doing science in language minority classrooms, to learn science or to learn English?
- Is there an alternative to common practice?
- What are the implications of such an alternative for assessment?

In this [illegible] we will explore these questions. Drawing on concrete examples of classroom science, we will elaborate an alternative to traditional practice which we refer to as scientific sense-making and discuss possible contexts and roles of assessment that emerge in a sense-making culture in language minority classrooms. We also explore the implications of this view for improving science education and assessment for language minority students, paying particular attention to issues of teacher development.

Why Do Science?

In bilingual programs this question looms large. In many cases, science is not taught at all. In those cases where it is a part of the curriculum, it is often seen as a context for learning English. Its intrinsic value as an academic discipline, as a way of thinking and knowing, is either ignored or not recognized.

As we have argued elsewhere (Warren, Rosebery & Conant, in press), a *pluralistic* view of language and literacy (cf. Literacies Institute, 1990) not only reframes the problem of what it means to learn science but helps us better understand the relationship between doing science and literacy development. According to this view, knowing a language entails knowing more than *the* English language or *the* Spanish language or any other language. Each language is really many languages, a set of possible discourses people use to communicate with one another in their daily activity (Bakhtin,1981). These discourses in turn each constitute a set of beliefs and values in terms of which one speaks, thinks and acts (Gee, 1989). The particular discourse worlds we inhabit depend on our history, the books we have read, the people with whom we have talked and from whom we have learned, the social circles in which we have moved, our economic class, our generation, our epoch, the institutions (church, political party, schools, societies) to which we have belonged, and so forth (Booth, 1986). As the Soviet theorist, Mikhail Bakhtin (1981:291), explains:

> At any given moment of its historical existence, language...is heteroglot from top to bottom: it represents the co-existence of socio-ideological contradictions between the present and the past, between differing epochs of the past, between different socio-ideological groups in the present, between tendencies, schools, circles and so forth, all given a bodily form. These "languages" of heteroglossia intersect each other in a variety of ways, forming new socially typifying "languages."

The idea that language is heteroglot poses some difficulties for both our common sense and technical uses of terms such as language (as in "learning the English language") and literacy. In both senses, these terms are often used to suggest a capability that is unitary and univocal rather than pluralistic and multivocal (although the varied definitions of literacy that abound in the literature are perhaps a clue to its inherent diversity). In the same vein, language and literacy often are defined in terms of mastery of certain general skills -- reading, writing, arithmetic skills -- rather than in terms of mastery of whole systems of meaning and practices, each involving a set of beliefs and values or, in Bakhtin's term, an *ideology*.

From within this sociocultural perspective on language and literacy, then, we do not view science as a context for developing English language skills. Nor do we define scientific literacy as the acquisition of specific knowledge ("facts") or general skills (e.g., observation, inference) or correct mental models. Rather we understand scientific literacy to be a socially and culturally produced way of thinking and knowing, with its own sense-making practices, its own values, norms, beliefs, and so forth. In this light, when students learn science, they are *appropriating* socially mediated ways of knowing, thinking, acting and using language (both first and second languages) to construct scientific meanings.

The task facing the second language learner -- and, specifically, in this culture, the learner of English -- is therefore enormously complex. Learning in school really means appropriating *whole systems of meaning* involved in such tasks as reading and answering questions about stories, talking to the teacher, taking tests, playing with other students in the school yard, doing mathematics, doing science, doing history, and so on (cf. Gee, 1989; Michaels & O'Connor, 1991). The notion of appropriation is key because it casts the learner as someone who is trying to find ways to take the sense-making practices of science, for example, and make them his or her own, tuning them to his or her own intention, his or her own sense-making purposes. As Bakhtin (1981:293-294) explains, appropriating a new discourse is a difficult process:

> (The word in language) becomes "one's own" only when the speaker populates it with his own intention, his own accent,

> when he appropriates the word, adapting it to his own semantic and expressive intention. Prior to this moment of appropriation, the word...exists in other people's mouths, in other people's contexts, serving other people's intentions: it is from there that one must take the word, and make it one's own. And not all words for just anyone submit equally easily to this appropriation, to this seizure and transformation into private property: many words stubbornly resist, others remain alien, sound foreign in the mouth of the one who appropriated them and who now speaks them; they cannot be assimilated into his context and fall out of it; it is as if they put themselves in quotation marks against the will of the speaker. Language is not a neutral medium that passes freely and easily into the private property of the speaker's intentions; it is populated -- overpopulated -- with the intentions of others. Expropriating it, forcing it to submit to one's own intentions and accents, is a difficult and complicated process.

What makes appropriation so difficult is that discourses are inherently ideological; they crucially involve a set of values and viewpoints in terms of which one speaks, acts, and thinks (Bakhtin, 1981; Gee, 1989). As a result, discourses are always in conflict with one another in their underlying assumptions and values, their ways of making sense, their viewpoints, the objects and concepts with which they are concerned. Each gives a different shape to experience. Therefore, appropriating any one discourse will be more or less difficult depending on the various other discourses in which students (and their teachers) participate. As Michaels & O'Connor (1991:11) explain,

> This conception of literacy has strong implications for how we think about cultural diversity and the knowledge that students bring with them from home. Each child in this society learns culturally appropriate ways of using language and of taking meaning from written texts in the early years at home. Cultural groups in this society have sophisticated ways of integrating the written language around them into their daily social life. However, ways of using oral and written language are closely tied to culturally different ways of interacting with others and with culturally different values and attitudes. Some children have home-based ways of using language that are more closely related to the ways in which language is used in schools than are the home-based practices of other children.

For language minority students, the appropriation process can therefore be more arduous than for other students, for the distance they must travel between discourse worlds -- ways of organizing an argument, interpreting questions -- is often far greater. As research has shown (Au, 1980; Au & Jordan, 1981; Michaels, 1981; Mohatt & Erickson, 1980; Philips, 1972), conflicts between school-based ways of using language and minority students' home-based practices can create barriers that limit minority students' access to the discourses that are needed to achieve in this society (Heath, 1983; Michaels & O'Connor, 1991).

Within the framework we are putting forward here, the key question then becomes: In what ways can language minority children be enculturated into the community of scientific discourse? What does it mean to do science? These are the questions to which we now turn our attention.

Science as a Sense-Making Practice

A new conceptualization of learning is emerging in the research literature (Brown & Campione, in press; Brown, Collins & Duguid, 1989; Lampert, 1990; Resnick, 1989; Schoenfeld, in press-a, in press-b). Drawing heavily on Vygotsky (1978, 1985) and on anthropological perspectives on learning and cognition (Geertz, 1973, 1983; Lave, 1988), this literature views learning as an inherently cognitive and social activity. The child appropriates new forms of discourse, knowledge, and reasoning through his or her participation in socially defined systems of activity. As Resnick (1989) has recently argued, education may be better thought of as a process of socialization, rather than instruction, into ways of thinking, knowing, valuing, and acting that are characteristic of a particular discipline.

Central to this view is the idea that concepts are constructed and understood in the context of a community or culture of practice; their meaning is socially constituted (Brown, et al., 1989). Within this community, moreover, practitioners are bound by complex, socially constructed webs of belief which help to define and give meaning to what they do (Geertz, 1983). As Mehan (in press) has noted, members of a community "cannot make up meanings in any old way." Rather, they build up ways of knowing, talking, acting, and valuing, which help to constrain the construction of meaning within the discipline. Within this framework, the learner is conceptualized as one who appropriates new forms of knowledge through apprenticeship in a community of practice (Brown & Campione, in press; Brown et al., 1989; Collins, Brown & Newman, 1989; Lampert, 1990; Lave, 1988; Resnick, 1989; Rosebery et al., 1990; Rosebery et al., in press; Schoenfeld, in press-a, in press-b; Warren et al., 1989).

What, then, is the nature of scientific practice? For the Nobel Laureate, scientist, Sir Peter Medawar (1987:129), scientific sense-making is a kind of storytelling:

> Like other exploratory processes, (the scientific method) can be resolved into a dialogue between fact and fancy, the actual and the possible; between what could be true and what is in fact the

> case. The purpose of scientific enquiry is not to compile an inventory of factual information, nor to build up a totalitarian world picture of Natural Laws in which every event that is not compulsory is forbidden. We should think of it rather as a logically articulated structure of justifiable beliefs about a Possible World -- a story which we invent and criticize and modify as we go along, so that it ends by being, as nearly as we can make it, a story about real life.

Medawar's use of the story metaphor represents a bold challenge to both typical school beliefs about what it means to be scientifically literate and the larger culture's assumptions about the nature of scientific knowledge. First, he challenges the belief that science, at bottom, is the discovery of a reality that exists "out there," pregiven but hitherto concealed (cf. Latour & Woolgar, 1986). Secondly, he challenges the belief that scientists work according to a rigorously defined, logical method, known popularly as The Scientific Method. And thirdly, through his emphasis on story building, he challenges the belief that scientific discourse -- the construction of scientific meaning -- is represented uniquely by forms of writing and talk that are thoroughly objective and impersonal.

Central to Medawar's vision is an idea of scientific practice in which creativity and construction -- rather than discovery -- predominate. His language suggests that science is projective rather than objective: scientists build stories about a *Possible* World, they do not discover the truth that already exists "out there." Further, he insists on the dialogic quality of scientific activity: fact and fancy, invention and criticism interacting.

Contemporary sociological and anthropological studies of the nature of scientific activity in laboratory settings add an explicit social dimension to this picture (Knorr-Cetina & Mulkay, 1983; Latour, 1987; Latour & Woolgar, 1986; Longino, 1990; Lynch, 1985). These studies show that scientists construct and refine their ideas within a community in which they transform their observations into findings through argumentation and persuasion, not simply through measurement and discovery. The apparent "logic" of scientific papers is really the end result of the practice of a group of scientists whose goal is to eliminate as many alternative interpretations as possible to their account of the phenomena being studied. Rather than the orderly, logical and coherent process that is described in science textbooks as The Scientific Method, actual scientific practice entails making sense out of frequently disorderly observations and negotiating among alternative interpretations. Through graphs, notes, statements, drafts of papers, and published papers, accounts are constructed, claims are negotiated, analogies are sought, arguments are put forward and defended against attack, and objections are anticipated (Latour & Woolgar, 1986). As Latour and Woolgar show, scientists claim merely to be discovering facts but close observation reveals that they are writers and readers in the business of being convinced and convincing others. (It is hard not to hear an echo of Medawar's storytelling in this.)

Through our work with bilingual teachers and students (Rosebery et al., in press; Warren et al., in press; Warren, Rosebery, Conant & Barnes, 1990), we are attempting to elaborate an approach to science teaching and learning that supports the development of scientific sense-making communities in the classroom. The basic idea is to create a community in which what the students think--the sense they are making of the world -- rather than what the text or teacher thinks is at the center of the class activity. This approach entails a radically different orientation to teaching and learning than that found in traditional classrooms, one in which students construct their scientific understanding through an iterative process of theory building, criticism and refinement organized around their own questions, ideas, and data analysis activities. Fundamentally, the idea is to place question posing, theorizing, and argumentation at the heart of students' scientific activity. Students explore the implications of the theories they hold, examine underlying assumptions, formulate and test hypotheses, develop evidence, negotiate conflicts in belief and evidence, argue alternative interpretations, provide warrants for conclusions, and the like. Conceptually, they investigate their own questions and the beliefs or theories from which they derive; epistemologically, they explore relationships among truth, evidence, and belief in science. They, in short, become authors of ideas and arguments (cf. Lampert, 1990).

In addition, students' inquiries are collaborative in nature, just as is most professional scientific activity. The emphasis on collaborative inquiry reflects our belief, building on Vygotsky (1978), that robust knowledge and understandings are socially constructed through talk, activity, and interaction around meaningful problems and tools. Collaborative inquiry provides direct cognitive and social support for the efforts of a group's individual members. Students share the responsibility for thinking and doing, distributing their intellectual activity so that the burden of managing the whole process does not fall to any one individual. The distribution and sharing of intellectual responsibility is particularly effective for language minority students, for whom the language demands of tasks are often overwhelming and can often mask their abilities and understanding. In addition, collaborative inquiry creates powerful contexts for constructing scientific meanings. In challenging one another's thoughts and beliefs, students must be explicit about their meanings; they must negotiate conflicts in belief or evidence; and they must share and synthesize their knowledge in order to achieve a common goal, if not a common understanding (Barnes & Todd, 1981; Brown & Palincsar, 1989; Hatano, 1981; Inagaki & Hatano, 1983).

Students' investigations are also interdisciplinary; science, mathematics and language use (talk, reading, and writing in both first and second languages) are intimately linked. Mathematics and language are recognized as essential tools of scientific sense-making, which stands in sharp contrast to traditional schooling in which science is separated from math and the role of language in each is hardly acknowledged (or, as in the case of many bilingual science programs, the relationship between science and language is reversed). The importance of an interdisciplinary approach cannot be overstated with regard to language minority students. It involves them directly in the kinds of purposeful, communicative interactions that promote genuine language use, which arguably are the most productive contexts for language acquisition, such as talking in the context of doing science and trying to solve a meaningful problem. It also creates opportunities for students to use the languages of science and mathematics in ways that schools and the society at large require: not just to read textbooks or do computations, but to write reports, argue a theory, develop evidence, and defend conclusions.

A brief example will help illustrate what we mean. In a Haitian bilingual combined seventh and eighth grade, students explored relationships among truth, belief and evidence in science through an investigation of their school's water. In the Water Taste Test, the students actively tested a widely held belief that the water from the school's third floor fountain was better than that from the other floors. With guidance from their teacher, they formulated their belief as a question and designed an investigation to explore its 'truth'. A blind taste test of about 40 of the school's junior high students revealed that most of them actually preferred water from the first floor although they believed they preferred water from the third floor. This finding prompted the class to pose a new question about the source of the difference and to investigate more deeply the physical and chemical quality of the school's water. Their analysis led them to conclude that temperature was a deciding factor in taste preference, but it also uncovered surprisingly high bacteria levels in the school's water. In the Water Taste Test, and possibly for the first time, the students themselves took control of their learning and, through scientific inquiry, constructed knowledge that was meaningful to them, their teacher, and the larger school community.

This discussion raises the question of the teacher's role in a sense-making culture. Far from backgrounding the teacher's function, a sense-making perspective on classroom practice intensifies it, as Duckworth (1986:133) explains:

> The essential condition of having the students do the explaining is not the withholding of all the teacher's own thoughts. It is, rather, that the teacher not consider herself/himself the final arbiter of what the learner should think, nor the creator of what

that learner does think. The important job for the teacher is to keep trying to find out what sense the students are making.

"Finding out what sense the students are making" of the phenomena they are exploring and, indeed, of their own thinking process entails a significantly different orientation to teaching, learning, and assessment than that found in most science classrooms. Above all, perhaps, it creates uncertainty by advancing a view of knowledge as a human product and a view of classroom discourse as a social process in which argument and conjecture play central roles (Cohen, 1988). This was the case in the Water Taste Test in which the teacher had no idea where the students' investigation would lead, what 'answer' it would produce. Sense-making also entails probing students' talk to find out how they are thinking, the assumptions they are making, the rationale behind their method. It includes helping students think through partial ideas or strategies in ways that do not undercut their own intentions, and involving other students in that process. It also includes valuing alternative interpretations and methods, helping students to explore the implications of their ideas and make connections between their own ways of thinking and scientific ways of knowing (cf. Lampert, 1990). To orchestrate these sense-making interactions, teachers must also have command of the domain knowledge involved in the students' inquiries. For the Water Taste Test, for example, the teacher and students learned about the chemistry of water quality analysis, aquatic ecosystems, hydrology, and water resource management. In a sense-making culture, therefore, "process" and "content" are inextricably linked; teachers guide students in making sense of real phenomena.

Contexts of Assessment in a Sense-Making Community

It may seem odd that in a paper on assessment we have dwelt so long on developing an image of a different kind of classroom scientific practice. But, in fact, this emphasis highlights a crucial point. In the current discussion on the need for accountability (U.S. Department of Education, 1991), there is a danger that we will neglect a critical question: Accountable for what? What is it that we want our children to learn? What does it mean for a student to be scientifically literate? We must not assume that because we are ready to reform assessment we fully understand the thing it is we want to assess better. For this reason, we have chosen to present scientific sense-making as a way to do science in order to anchor our discussion of assessment in a particular context and to emphasize the importance of taking into account the local -- as opposed to the national -- character of assessment.

As outlined in the previous section, scientific sense-making reconfigures teaching and learning in some significant ways. Unlike conventional classrooms, teaching and learning in such a culture are not bound to textbooks, canonical experiments with their correct outcomes, or even a curricular scope and sequence. Students pose questions, design research, use tools to make sense of the world, collect data, build and argue theories, and document and communicate their findings and interpretations in various ways. Students' inquiries stretch over long periods of time, not just weeks but in many cases months. They take unexpected turns. The context of students' scientific work is social rather than individual. Further, in a sense-making culture, teachers take on a variety of roles; they coach and model scientific practices, and they act as co-investigators.

Given this radical change in the classroom culture, in the kinds of processes and products that characterize learning, our concern in this section is to explore some possible contexts of assessment that are congruent with sense-making and that tap the full range of students' learning. In particular, we explore the varieties of learning (Michaels & O'Connor, 1991) that are made manifest through students' talk and writing as they construct scientific meanings. The examples are drawn from classrooms that are working to establish sense-making communities in science. We have chosen examples from a collaboration involving two Kindergartens -- one Haitian bilingual and one English monolingual -- and a multilingual/multicultural basic skills high school class to show that the kinds of activity and reasoning that emerge in a sense-making culture are as appropriate for five-year-olds as they are for sixteen-year-olds. In the concluding section, we explore how the role of assessment in a sense-making culture can be extended beyond student monitoring to promoting learning and teacher reflection.

Students' Talk: Examples from a Kindergarten Collaboration

Talk is highly valued in a sense-making community as a means for negotiating and constructing scientific meanings. Through talk, students make their thinking public, argue alternative theories, collect data, elicit assumptions, pose questions and conjectures, among other things. Classroom talk falls on a continuum; at one end, it can be organized as a teacher-moderated classroom discussion and at the other it can be spontaneous as in an informal conversation between students analyzing data at a computer. It is, in short, socially situated and multidimensional.

Research has shown that classroom discourse in various domains is enormously complex (Adelman, 1981; Barnes, 1976; Cazden, 1988; Cazden, John & Hymes, 1972; Cook-Gumperz, 1986; Edwards &

Mercer, 1987; Michaels, 1981; Heath; 1983; Wells, 1981). Recent studies have begun to explore the discursive, linguistic, and cognitive characteristics of science and mathematics and the relationship between student learning and the ways in which teachers orchestrate talk (Lampert, 1990; Lemke, 1990; Michaels & Bruce, 1989; Michaels & O'Connor, 1991; Rosebery et al., in press). In at least one case, classroom discussion in science has been explored as a context for assessment of students' thinking (Chittenden, 1990). Taken together, these studies suggest that talk represents a rich, but challenging, context for learning about how students are making sense of the world.

In the following, we explore several examples of classroom talk from a collaborative weather investigation conducted by two Kindergartens, one Haitian Creole bilingual, the other English monolingual. The teachers of these classrooms informally observed and monitored their students' daily use of scientific tools (e.g., thermometers, wind socks, anemometers, rain gauges, bargraphs and charts for representing data) and their talk as a basis for assessing their learning. Our focus is on the varieties of learning that emerge from an analysis of talk.

For the better part of the school year, a Haitian Creole bilingual Kindergarten and a monolingual (English) Kindergarten collaborated on an in-depth investigation of their local weather. Students investigated and collected data on clouds, wind direction and speed, precipitation, and temperature to explore their influence on local weather patterns. They learned to use an anemometer to calculate wind speed, and wind socks and a stationary compass painted onto their school playground to determine wind direction. They observed clouds, noting their color, formation, number, approximate height, and movement; based on these observations, they invented a taxonomy of cloud types. They also learned to use a thermometer to determine air temperature and to check the accuracy of their daily temperature predictions (which became increasingly accurate as the investigation progressed).

Each day, small groups of students collected and recorded data, represented those data in graphs, composed stories of their observations, and the like. They also worked in large groups, reporting their data and observations to one another, and asking and answering each others' questions. Some of their questions included: "What makes the clouds change so quickly?" "Why does the wind sock blow one way and the clouds go another?" "What makes the wind?" "Does it always get colder when it rains?" In the spring, the classes met together on a daily basis to report and discuss their findings and to examine their data (wind speed and direction, temperature, precipitation, cloud cover) for interesting patterns and relationships.

In the following examples of classroom talk, students demonstrate both scientific and mathematical reasoning through tool use and data analysis. The first two examples are taken from a class held in the Spring of 1990 in which three Haitian Kindergartners are reporting the day's weather data to an audience comprised of both bilingual and monolingual Kindergartners and their teachers. In Example 1, Georges, Jonese, and Frantzia are being prompted by their teacher, Christine, to report on wind direction. The exchange takes place in English.

Example 1

Christine: What about the wind? Where was the wind blowing to?
Georges: FROM, Christine!
Christine (smiles and laughs): FROM!...Where was the wind blowing FROM?
Georges, Jonese, Frantzia: From east to north.
Christine: How did we find that out? What did we use for that?
Jonese: The wind socks.
Christine: And where did we stand?
Jonese: In the middle...of the...the compass.

The focus of this exchange is the reporting of the day's wind direction. The most remarkable aspect of the exchange, which lasts under a minute, is the opening two lines when Georges notes aloud that the teacher has misspoken, saying "...to..." instead of "...from..." in talking about wind direction. In correcting Christine, Georges demonstrates that he has not only learned to use the wind sock to determine wind direction but has also learned the standard meteorological convention for reporting it. This standard, incidentally, is not intuitive and is easily confused by adults (as Christine demonstrates). A wind sock, for example, shows very clearly the direction *to* which the wind is blowing; determining the direction *from* which it is blowing requires an inference. Georges's two word counter to Christine makes clear that he has learned how to *talk and think* about wind direction, and that he is not afraid to assert this knowledge. That he insists on maintaining the convention they have established through their own field work is also evidence of the value he places on their work. As we noted earlier, citing Mehan (in press), a discourse community collectively builds up its ways of talking and knowing; it doesn't "make up meanings in any old way." Georges's concern for maintaining the classroom community's standard does not go unnoticed by Christine who laughs good naturedly at his correction and then takes it up in her restatement of the question.

Example 2 takes place a few minutes later in the class. It is an exchange involving both bilingual and mainstream students under the guidance of Christine, the bilingual teacher. Georges, Jonese, and Frantzia have finished reporting their weather data, which included discussion of two different readings obtained for wind speed. In the front of the school they observed that the anemometer made one revolution which calculated to a wind speed of zero miles per hour; in the back of the school (specifically in the teachers' parking lot on top of Christine's car), they observed it make two revolutions which calculated to a wind speed of one mile per hour. At the point we join the conversation, Christine has invited the students to ask questions (a standard practice) and Johnny, a bilingual student, asks Georges why they got two different wind speed readings. Later in the conversation, Susan, a monolingual student, joins in. The exchange takes place in English.

Example 2

Johnny: To Georges, why when you put the anemometer on Christine's car did it turn but not in front?

Christine: That's a good question. Johnny said, "Why when we put the anemometer on the top of the car we had two revolutions and when we had it in front there was only one revolution?" Who can answer that question?

Jonese: Me!

Christine: Jonese? Ok, Georges wants to try it first because it was to Georges.

Georges: (inaud)...

Christine: Ok, Jonese wants to try it.

Jonese: I think that when it was in front it didn't have no wind and when we were in the back and put it on the top of the car it was a little windy and cold and we had two revolutions; first there was one in the front; then there were two.

Christine: WHY is it more windy in the back than in the front? Susan wants to try that.

Susan: Maybe because the building keeps the wind from going around to the front.

Christine: OKAY! Who else has a question. Susan?

Susan: I have a question for you.

Christine: Me!? I hope I can answer it!

Susan: Since it went around two times, does it always go one less miles per hour on the computer?

Christine: Mmmhumm, when it goes around two times, it's one mph, but when it goes around one, it's zero.

In this exchange the students' reasoning is striking. First, Johnny shows that he is thinking critically about the data that have been presented. He articulates what he feels is an inconsistency in the data and demands an explanation. This is exactly the kind of scientific thinking the teachers have been trying to promote in their students throughout the year. The discourse context is not simply

Show and Tell but a forum for making sense of the data the students have generated through their own scientific activity. In this context, then, making sense of an inconsistency in data is standard practice.

Secondly, the explanations that Jonese and Susan generate provide information about each girl's control over the discourse of scientific explanation, at least on this day and in this situation. Their explanations differ in crucial ways. Most significantly, Jonese's response does not meet the (implicit) criterion the teachers have established for scientific explanations. She offers a reason for the difference in wind speed, saying that in front there wasn't any wind while in back there was, but, according to the teachers' standard, it is tautological; it doesn't explain the data so much as repeat them. Christine notes this in her response by rephrasing Johnny's question to emphasize causation ("WHY is it more windy in the back than in the front?"). Susan's response, in contrast, is closer to the teachers' notion of explanation; it contains an explicit marker ("because") and elaborates a plausible reason ("Maybe because the building keeps the wind from going around to the front."). This example raises an important question. While Jonese and Susan respond very differently to the call for an explanation, the talk itself does not help us understand *why*. Is it because Jonese is less familiar with the discourse of explanation in this context? Does she not understand the teacher's question and its implicit discourse assumptions (Michaels & O'Connor, 1991)? Are the criteria for explanations too implicit (Delpit, 1986, 1988)? Regardless, Jonese's difficulty should serve as a signal to her teacher to probe its source more deeply, or in Duckworth's words, "to find out what sense the students are making." We will return to this example in the next section when we discuss the role of assessment in a sense-making culture.

A third and final snapshot of students' learning in this exchange is represented in Susan's question to Christine ("Since it went around two times, does it always go one less miles per hour on the computer?"). This question is noteworthy for what it reflects about the depth and nature of Susan's thinking. On the basis of the data presented in class that morning, Susan poses a question to test a rule for calculating the wind speed in miles per hour based on the number of revolutions of the anemometer (something like: wind speed = number of revolutions – 1). While her algorithm is not correct, it is evidence that she is examining the data for patterns and then using those patterns as the basis for generating rules, a highly sophisticated form of reasoning. At the time of the exchange, neither Christine nor the monolingual teacher understood that Susan was testing a generalization. Prompted by one of the researchers, Christine follows up with Susan the next day (unfortunately their conversation was not recorded). Christine reported afterward that once she understood Susan's intended meaning, they went to the cumulative weather chart the classes had been developing and together examined several days' worth of wind speed data (revolutions and mph). In this way, Susan discovered for herself that her rule was not supported by the data. This has implications for assessment. By asking Susan to join her at the chart to evaluate the rule against the data, Christine is helping Susan to answer her own question and in the process is introducing her to a standard scientific practice for evaluating a rule or conjecture. Moreover, by scaffolding Susan's activity, she is enabling her to accomplish more than she could have done on her own (Palincsar & Brown, 1984). Although Susan's rule was disconfirmed, her impulse to build generalizations based on observed patterns was shaped, extended, and reinforced through the teacher's action.

By way of closing our discussion of students' talk as a context for assessment we will examine an exchange that took place between three boys in the bilingual Kindergarten, Johnny, Pierre, and Josef, in an informal interview situation. The boys are being asked to read and interpret a set of daily temperature graphs (barcharts) their class has developed over several months. The discussion takes place in Haitian Creole and appears below in Haitian Creole followed by English translation.

Example 3

Interviewer: Ki sa ki deye nou la? (They turn.)
Johnny: Yon bagay ki pou weather a.
Josef: Le-l fe cho oubyen fret. Le bagay la ba, se cho l ap fe.
Interviewer: Se vre?
Johnny: No, fret!
Josef: Yeah—Le l wo se
Johnny and Josef: *cho*!
Interviewer: OK, ou ka gade sou premye a, sa ki an le a, ou ka di nou ki jou ki te fe pi cho an janvye? Ki jou ki te fe pi cho?
Johnny: Saa? (Pointing to the highest bar in the middle of the graph.)
Interviewer: Ki nimewo ou we li ba ou?
Johnny: Yon sis avek yon kat.
Interviewer: OK—
Josef: Men ni, men ni men ni! (Pointing to the highest bar at the end of the graph, also with a value of 64.)
Interviewer: Konben li fe, Josef?
Josef: Yon sis avek yon kat.
Pierre: Mwen we *sa* ki cho (pointing to a day when the temperature was zero and another when it was around 30).
Interviewer: Se sa ki pi cho?
Pierre: Yeah.
Johnny: Sa ki pi fret la, se saa (pointing to the lowest bar on the graph).

Pierre: Pi *cho*, sa ki pi *cho*, sa ki *fret* la (pointing, it seems, randomly at bars on the graph).
Interviewer: Sa ki fe ou di sa? Sa ki fe ou konn se sa ki pi cho?
Scott: Paske li menm ki pi gwo pase saa, pase, sa pi gwo pase...(tracing with his finger up the side of the graph).
Interviewer: Johnny, ou ka ede-l ? Li di se premye a ki fe pi cho, eske se vre?
Johnny: (Shakes head "No".) Saa ki pi cho (pointing to the highest bar at the end).
Pierre: Sa ki gwo pase a (pointing to the highest bar in the middle).
Interviewer: Ou ka explike l poukisa, ou ka di Pierre poukisa?
Johnny: Se paske sa pi wo, li gen pi plis.
Pierre: Sa pa pi gwo (pointing to the highest bar at the end).
Josef: Li gen karant kat.
Pierre: Se paske sa (pointing to the highest bar in the middle) ki pi gwo pase saa (pointing to the highest bar on the end), epi sa pa pi gwo (pointing to the low bar next to the highest one on the end) sa (pointing to the highest bar in the middle) ki pi gwo pase a.
Interviewer: Sa ou panse, Josef? Kiles ki pi cho?
Josef: Saa ki pi wo (pointing to the highest bar at the end).
Interviewer: What's that behind you? (They turn.)
Johnny: A thing for the weather.
Josef: When it's cold or hot. When the thing is low, then it's hot.
Interviewer: Is that true?
Johnny: No, cold!
Josef: Yeah, when it's high it's —
Johnny and Josef: hot!
Interviewer: OK, can you look at the first one, the top one? Can you tell me which is the hottest day in January? Which day is the hottest?
Johnny: This? (Pointing to the highest bar in the middle of the graph).
Interviewer: What number is it?
Johnny: A six and a four.
Interviewer: OK—
Josef: Here it is! Here it is! Here it is! (Pointing to the highest bar at the end of the graph, also with the value of sixty-four.)
Interviewer: How many is it, Josef?
Josef: A six and a four.
Pierre: I see *this* is hot (pointing to a day when the temperature was zero and another when it was around 30).
Interviewer: That's the hottest?
Pierre: Yeah.
Johnny: This is the *coldest* one (pointing to the lowest bar on the graph).
Pierre: *Hotter*, this one's hotter, this one's cold (pointing, it seems, randomly at bars on the graph).

Interviewer: Why do you say that? How do you know it's hotter?
Pierre: Because this is bigger than this, this is bigger than this (tracing with his finger up the side of the graph).
Interviewer: Can you help him, Johnny? He says that the first one's hotter, is that true?
Johnny: (Shakes head "No".) This is the higher one (pointing to the highest bar at the end).
Pierre: This one's higher than it (pointing to the highest bar in the middle).
Interviewer: Can you explain to Pierre why?
Johnny: Because this is the tallest, it has the most.
Pierre: That's not the biggest.
Josef: It's forty-four!
Pierre: It's because this one (pointing to the highest bar in the middle) is bigger than this one (pointing to the highest bar on the end), and this one (pointing to the low bar next to the highest one on the end) isn't big[-ger than] the one bigger than it.
Interviewer: Is that what you think Josef? Which is the hottest?
Josef: This one is the highest (pointing to the highest bar at the end).

While this discussion took place in an informal interview, we observed similar conversations taking place spontaneously as the children examined their graphs and data charts. It is clear from the above discussion that Johnny knows how to read and interpret a bar graph, relate it to the phenomena it represents ("This is the coldest one."), and articulate its meaning to others. Pierre, on the other hand, does not seem to understand the graph and, perhaps most distressing from a teacher's perspective, seems unaware of his own confusion. The state of Joseph's understanding is somewhat less clear from this bit of transcript. At the end, however, when he explains in response to the interviewer's question "Which is the hottest?" that the hottest day is the highest bar suggests that he does understand how the graph represents hot and cold temperatures, and that he can translate between different ways of making sense of the graph.

Our purpose in presenting the above examples is to demonstrate the richness of classroom talk and its relationship to student learning. Through their talk, students showed varieties of sense-making. They mastered the use of specific tools and the concepts underlying their use; they interpreted graphs, critically analyzed numerical data and suggested generalizations based on those data; they built explanations and posed questions focused on data they had generated. In addition, the focus on classroom discourse brought to light instances of talk in which the meaning of that talk was not understood, either by the teacher or the student. These instances underscore the need for explicit discussion of the standards and assumptions for talk in a scientific community (Delpit, 1986; 1988; Michaels & O'Connor, 1991). But this suggestion should not be construed as a call for teaching students the 'rules' of talk or specific forms of explanation or vocabulary. Rather, it means that the classroom community itself needs to reflect on its talk in order to establish its own standards and uncover implicit assumptions. Concern for talk -- how to put forward effective arguments, pose provocative questions, and marshall convincing evidence -- should become an integral part of the work teachers and students see themselves doing, that is, a distinguishing feature of their work as members of a scientific sense-making community (cf. Brown & Campione, in press).

Students' Writing: Examples from a High School Field Ecology Study

Portfolios represent one variation on the theme of alternative assessment in writing, one that also has potential in science. As Wolf, Bixby, Glenn & Gardner (in press) explain, the concept of a portfolio itself has begun to evolve from a structured sampling of a student's work over time to the idea of a process-folio (Gardner, 1989, in press; Wolf, 1990) which differs in several ways from the traditional conception:

> [Process-folios] differ from familiar portfolios in a number of ways. The generation of these process-folios is embedded in a much larger classroom context where teachers and students frequently discuss what goes into creating worthwhile work, what makes for helpful critique, and how to plow comments back into ongoing work. In addition to finished works, these collections contain sample "biographies of work" -- documentation of the various stages of a project. When collected at diverse points, these biographies permit a longitudinal look at a student's changing control of the processes for shaping a final piece. Students often keep journals and write reflections about their work (Seidel & Zessoules, 1990). Finally, the collections of work students build are anything but archival. They regularly return to earlier works to revise or make comparisons with later ones. At the close of the year, students reenter their collections to make a final selection of biographies, reflections, and final pieces that can serve as the basis for a course grade and/or part of a permanent record of their development (Camp, 1990a, 1990b; Howard, 1990; Wolf, 1989). In this sort of work, students have the opportunity to see samples of different levels of work and to discuss the criteria that distinguish strong performances. They also witness the multidimensional nature of such work (i.e., that it involves the ability to pose an interesting problem, to learn from and comment on someone else's work, or to revise an earlier draft.) (Wolf et al., in press:34)

The idea of process-folios strengthens the link among teaching, learning, and assessment by blurring the boundaries which in conventional practice separate them. Process-folios represent an intriguing possibility not only for capturing the complexity and richness of students' scientific sense-making but for making assessment a more integral part of what teachers and students see themselves as doing in the classroom. In a sense-making culture, students' work is not only sustained over long periods of time but is subject to critique, review, false starts, new questions, and a variety of choices that are often contextually contingent. As students conduct investigations, they keep notebooks that contain a wide range of informal "writing" including questions, hypotheses, data tables, graphs, notes about experimental procedures, informal analyses and interpretations of data, and the like. They also produce formal texts such as charts, graphs and reports for publication, i.e., for an outside audience.

In this section, we look at examples of the informal scientific writing of two Haitian students, Rose and Marie. We analyze their texts for evidence of the ways in which they are making sense of data they developed. Both students were in a multilingual basic skills class in a large urban high school. (Six different languages were spoken in the class.) Their class was composed of students who were judged not ready for the regular bilingual program because of low academic skills. For the most part, these students could not read or write their first language or English.

During the school year, the class studied water quality using their home tap water as the basis of study. In the spring, their interests broadened to encompass an ecological study of a local pond that bordered the city's water reservoir. The students were concerned that the pond, which was obviously polluted, posed a threat to the city's drinking water. To address their concern, the students decided to study the health of the pond, including an analysis of its chemical, biological, and physical characteristics, and to investigate the city's water supply, learning about its sources, how it is purified, and how it is piped throughout the city. To complete their investigation, the students broke into small groups to work on particular aspects of the study.

Rose's group, for example, was responsible for determining the bacteria level of the pond. In keeping with their year-long interest in home water, the group decided to compare the bacteria level of the pond to that of their local drinking water. They were interested in two things: How much bacteria was in the pond? How much bacteria was in their drinking water? To answer their questions, they collected water samples from the pond, their homes, and school drinking fountains and tested them.

To perform this test, the students used commercially available culture kits called Millipore™ samplers. These samplers are made of an absorbent, nutrient-filled pad which is marked with a grid. To test for bacteria, the pad is immersed in a water sample, placed inside a plastic container, and incubated under a lamp for twenty-four hours. At the end of twenty-four hours, the grid is inspected for bacteria colonies which appear as tiny black, blue, or green spots. A pamphlet accompanying the samplers allows the user to assign a water quality grade based on the number of colonies that grow. To be drinkable, water must have a count of zero.

For undetermined reasons, many of the students' cultures did not grow. A few did, however, and Rose used them as the basis for investigating the bacteria level in the city's tap water. Her first step was to document her results. She drew a facsimile of the Millipore™ sampler in her lab notebook, reproducing the position and size of each of the 57 bacteria colonies that had grown. Her drawing was a meticulous and accurate reproduction of the culture. She then interpreted the significance of her findings. According to the standards stated in the Millipore™ pamphlet, the tap water, which had come from a student's home, was not fit to drink. Rose documented her findings in her notebook in English as follows:

> I counted the bacteria in the tape water.
> I find fivety seven bacteria in the tape
> water. That's mine you can't not drinking
> but you can swim on that water --
> Grade B for that water because whole body
> contact no more than 200/100 ml.

Rose's report, brief as it is, draws on diverse resources and voices to communicate her finding and its significance. For example, because she is concerned that her report be viewed as credible within her scientific community, she uses two devices common in the discipline to lend her argument validity -- referring to other literature and making her data publicly available. She establishes a connection, if only implicitly, with the standards that accompany the Millipore™ samplers ("That's mine you can't not drinking but you can swim on that water -- Grade B for that water because whole body contact no more than 200/100 ml."). To lend a sense of precision and verifiability to her report, she includes her representation of the sampler and reports the bacteria count in her analysis, in this way documenting her interpretation.

The discourse strategies Rose uses to organize her report also reflect her desire to communicate her scientific activity in accurate detail. She describes how she came to her results and what she found, clearly marking them as the product of her own efforts through use of the first person authorial voice ("I counted..." "I find..."). Not only is she reporting her scientific method but, by using the first person, she marks her result as a personal construction which does not exist apart from her work or reasoning. Note, however, that when she interprets the data according to the standards, Rose switches from the first person to the more authoritative, objective voice signalled in: "That's mine (That means) you can't not drinking but you can swim on that water. Grade B for that water because whole body contact no more than 200/100 ml." Here she is appropriating the words of the Millipore™ pamphlet to interpret her finding and to inform others of its significance: the water used in this sample is fit for whole body contact but not for drinking. (Grade B water, which is suitable for whole body contact such as swimming, can contain a bacterial count of 1-200 colonies per 100 ml of water.) This switch in voice suggests that Rose is aware that scientific results are reported "objectively," apart from the agent who produced them. The presence of both personal and objective statements in her report reflect her struggle to coordinate these voices as part of a coherent whole.

From an assessment perspective, what stands out in Rose's work is the way in which she takes control of the bacteria study, shaping it to her own purposes, taking a point of view, and then interpreting her activity and its significance for a larger community. Rose's activity and her report are evidence that she is beginning to think, act, and write like a scientist. The mixed levels of description and explanation, the orchestration of multiple voices, the recourse to standards and multiple representations reflect her own efforts at sense-making and belie the surface simplicity of her report. These sense-making efforts reflect her struggle to appropriate scientific ways of thinking, knowing, and writing. She is working through for herself the relationship between the processes by which she produced her finding ("I counted..." "I find...") and the means for communicating that finding ("That's mine..."). This effort is a key aspect of scientific practice, one that is well-known to anyone who has struggled to craft a "story" about data. That Rose does this in English, by her own choice, only adds to the complexity of her task. From a sense-making perspective, then, Rose's report, which on the surface seems simplistic and full of errors, is actually a complex text that shows she is beginning to forge a scientific voice.

About the time that Rose was finishing her study, the Basic Skills class was preparing for a field trip to the city's water treatment facility. The trip was set up so that the students would be able to ask questions of the city's water chemist at the end of their tour. In anticipation of this, the students were asked to generate questions they wanted to ask the chemist.

Many students had just finished reading a booklet, "The Story of Water," prepared by the city's Water Department which explains in pictures and words the water cycle and water treatment process. Under the direction of a classroom teacher, many of the students had developed a set of questions based on their reading. A quick survey of students' notebooks showed that approximately two-thirds of them contained the following kinds of questions:

> "What machines are used to purify water?"
> "What is chlorination?"
> "What is filtration?"

From a sense-making perspective, these questions are odd. They are about science content without being linked to authentic inquiry. They seek knowledge that is already known rather than knowledge that needs to be constructed. In short, there is little sense-making to be found in them. They are, however, typical of the kinds of questions students are frequently asked in school where the focus is on factual comprehension and recall.

In contrast, Rose and her partner, Marie, used the bacteria results as the basis for developing a different set of questions that grew directly out of their own scientific activity. The students first composed the questions that follow in Haitian Creole and then translated them into English:

> "I went a (sic) know how come bacteria come in the water?"
> "How come they clean (sic) the water but it still has bacteria in it?"
> "I went to know how often they clean the water?"

It is interesting to explore from an assessment perspective how Rose and Marie's questions differ from those of the rest of the class, and what they tell us about the students' scientific reasoning. As we noted earlier, the questions taken from the Water Department booklet have little to do with the students' own sense-making. "(I)t is as if they put themselves in quotation marks against the will of the speaker" (Bakhtin, 1981:293-94). The lack of student agency and purpose is perhaps most clearly reflected in the impersonal, objective voice in which the questions are cast. There is no sense of ownership, of the students actively asking and answering questions.

Rose and Marie's questions, however, presume an active, critical stance toward the world and, in particular, toward their finding. In a very real sense, their questions represent an action and assert a will to know ("I went to know..."). They literally call into question the dilemma posed by Rose's findings ("How come they clean the water but it still has bacteria in it?") and seek to resolve it. Unlike the class questions, these questions are openly purposeful and evaluative, expressing a particular point of view and designed to produce knowledge. Through their questions, Rose and Marie continue the process of sense-making initiated by Rose. Thus, while at first glance the class questions seem to be scientific in content and tone because they are "objective," they are not. In contrast, Rose and Marie's questions, which are markedly "subjective," are solidly grounded in scientific activity, evidence, and reasoning.

Writing of the kind presented above represents only one aspect of the work students produce in their scientific investigations. In addition, they write notes and make drawings of their observations, tabulate and represent data, design data collection instruments, and draft and finalize reports, among other activities. These texts are sometimes the work of an individual and sometimes the work of a group. They may represent half-baked ideas, rejected plans, or revised thinking. Thus, their role and use in assessing student learning needs to be carefully thought through. In fact, as Wolf et al. (in press: 27-28) suggest, such "assessment is not a matter for outside experts to design, rather it is an episode in which students and teachers might learn, through reflection and debate, about the standards of good work and the rules of evidence."

Roles of Assessment in a Sense-Making Culture

With the emphasis on performance, portfolios, and exhibitions, the assessment reform effort is attempting to blur the edges separating learning, teaching and assessment (Gardner, in press; Hein, 1990; Sizer, 1984; Wolf, 1989). These kinds of alternative assessments acknowledge the situated nature of cognition as they seek to explore student learning in complex, multidimensional activities that are representative of the work of a particular discipline. In some cases, they recognize both students and teachers as active participants in the process who set the standards to be applied to their work (Stock, 1990; Wolf, in press). In this atmosphere of critical reform, it becomes possible to rethink not only the means of assessment but also the roles it can play in teaching and learning. In this section we explore the implications of our prior analysis of student talk and writing for uses of assessment in the science classroom, particularly in promoting student learning and teacher reflection.

In the preceding section, we noted a difference in the kinds of explanations Jonese and Susan put forward for the wind speed data. We also commented that based on the talk itself it was impossible to determine *why* Jonese responded in the way she did. Moments like these represent one of the strongest arguments for linking teaching, learning and assessment as part of a larger enculturation process. It would be easy to judge Jonese as not having a theory to account for the difference in wind speed readings whereas Susan does. But her talk doesn't allow that inference. It is unclear from what she says whether she doesn't have a theory or doesn't understand the discourse assumptions underlying the teacher's question. In moments like these, it becomes the teacher's job to find out the reason and then to build on this knowledge to promote Jonese's learning, to help her gain access to the assumptions and rules that govern discourse in science in that classroom.

Teaching of this kind calls for a level of reflective practice (Schön, 1983, 1991) that is not only rare in schools (largely because it is not valued) but is also likely to be difficult to achieve without considerable effort and dedication of resources. But the benefits far outweigh the costs. To be convinced, we have only to consider the subsequent episode when Christine, after initially misunderstanding Susan -- and, as a result, missing the real import of her question -- returns to it the next day to find out her intended meaning which they then test against the evidence. Not only does the teacher learn something important about the depth of her student's reasoning but, by her action, she also places a high intellectual and psychological value on that reasoning. Taking students' questions seriously, probing their intended meaning, working to understand the assumptions on which they are based represent the kinds of actions that make teaching and assessment part of a larger reflective practice.

Not only do teachers need to become more aware of the complexity of classroom talk, writing, and activity and their relation to higher order thinking and discourse appropriation, as Michaels & O'Connor (1991) suggest, but they also need to develop better articulated views of science as a sense-making practice. These deeper understandings are needed if the effort to develop new forms of assessment is to succeed. For example, the scientific value of Rose's text and Rose and Marie's questions is not transparent. Indeed, it would be easy to be misled by the surface features of the texts (grammar, spelling, brevity) into underrating their scientific merit and the work that went into them. To appreciate the character of their sense-making requires having an insider's view of what it means to do science. This implies that teachers must become sense-makers themselves, as doers of science, teachers, and researchers interested in understanding and amplifying their students' ways of knowing. Indeed, in expert practice, these three roles interact; the ideal is a teacher who embodies and enacts all three as part of his or her classroom practice (Duckworth, 1986; Schön, 1983).

Helping teachers to think more deeply about science and classroom talk does not mean simply teaching teachers about new curricula or new teaching strategies. As a vehicle for change, innovative curricula are not enough; nor is current in-service (or preservice) education. While these may provide teachers with a grounding in the underlying scientific concepts and with hands-on activities to use in the classroom -- and, in the case of some in-service courses, with direct experience using those activities -- they do not touch on the deeper issue on which teacher change depends, namely, teachers' views of science and science pedagogy. The real issue is epistemological change, to bring about a shift in teachers' beliefs about science and pedagogy as well as a shift in their teaching practices toward a sense-making perspective.

Attempts to redraw the face of assessment in science must therefore be grounded in teacher development. The standards of good scientific practice cannot be imposed from outside the teaching community; they must be constructed from within, ideally through active debate not just between teachers and researchers but also between teachers and their students (Wolf et al., in press). In these ways assessment can become an occasion for both improving teaching and amplifying students' learning. Issues like these must be addressed as part of our reconceptualization of science assessment.

The significance of the links connecting teaching, learning, and assessment should not be underestimated. The analyses of student talk and writing we presented earlier represent *our* interpretation of students' scientific thinking based on our own view of what it means to be scientifically literate. The teachers' assessments were more informal, less tied to a view of science as sense-making. They tended to focus more on conventional categories such as students' facility with numbers, growth in language, quantity of talk, although as the year progressed they placed more value on the quality of the students' questions, their understanding of data, their critical-mindedness, and their initiative in defining questions or problems to explore. Their thinking on these issues continues to evolve. This, we believe, is where the hard work of assessment resides, in translating a view of what it means to be scientifically literate into criteria that can capture diverse student performances and varieties of thinking. This translation, moreover, cannot be made for teachers; rather it must be made *by* teachers based on their own elaborated understanding of what it means to be scientifically literate.

References

Adelman, C. (Ed.) (1981). Uttering, muttering: Collecting, using and reporting talk for social and educational research. London: Grant McIntyre.

American Association for the Advancement of Science (1989). Science for all Americans. Washington: Author.

Au, K. (1980.). Participation structures in a reading lesson with Hawaiian children: Analysis of a culturally appropriate instructional event. Anthropology and Education Quarterly, 11(2), 91-115.

Au, K. and Jordan, C. (1981). Teaching reading to Hawaiian children: Finding a culturally appropriate solution. In H.T. Trueba, G.P Guthrie, & K. Au (Eds.), Culture and the bilingual classroom: Studies in classroom ethnography (pp. 139-52). Rowley, MA: Newbury House Publishers.

Bakhtin, M. (1981). The dialogic imagination. Austin: University of Texas.

Barnes, D. (1976). From communication to curriculum. London: Penguin.

Barnes, D., and Todd, F. (1981). Talk in small learning groups: Analysis of strategies. In C. Adelman (Ed.), Uttering, muttering: Collecting, using and reporting talk for social and educational research (pp.69-77). London: Grant McIntyre.

Booth, W. (1986). Freedom of interpretation: Bakhtin and the challenge of feminist criticism. In G. Morson (Ed.), Bakhtin: Essays and dialogues on his work (pp. 145-176). Chicago: University of Chicago Press.

Brown, A. and Palincsar, A.M. (1989). Guided, cooperative learning and individual knowledge acquisition. In L.B. Resnick (Ed.), Cognition and instruction: Issues and agendas (pp. 393-451). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. & Campione, J. (in press.) Communities of learning and thinking or a context by any other name. Human Development.

Brown, J.P., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18 (1), 32-42.

Camp, R. (1990a). Arts Propel: Ideas for stimulating reflection. Unpublished manuscript. Princeton: ETS. Cited in Wolf et al. (in press).

Camp, R. (1990b). Thinking together about portfolios. The Quarterly of the National Writing Project, 27, 8-14.

Cazden, C.B. (1988). Classroom discourse: The language of teaching and learning. Portsmouth, NH: Heinemann.

Cazden, C.B., John, V. and Hymes, D. (1972). Functions of language in the classroom. New York: Teachers College Press.

Chittenden, E. (1990). Young children's discussions of science topics. In G. Hein (Ed.), The assessment of hands-on elementary science programs (pp. 220-247). Grand Forks, ND: Center for Teaching and Learning, University of North Dakota.

Cohen, D. (1988). Teaching practice: Plus a change... (Issue paper 88-3). East Lansing: Michigan State University, National Center for Research on Teacher Education.

Collins, A. Brown, J. S. & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L.B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 453-494). Hillsdale, NJ: Erlbaum.

Cook-Gumperz, J. (1986). The social construction of literacy. Cambridge: Cambridge University Press.

Delpit, L. (1986). Skills and other dilemmas of a progressive Black educator. Harvard Educational Review, 56(4), 379-385.

Delpit, L. (1988). The silenced dialogue: Power and pedagogy in educating other people's children. Harvard Educational Review, 58(3), 280-298.

Duckworth, E. (1986). The having of wonderful ideas and other essays on teaching and learning. NY: Teachers College Press.

Edwards, D. and Mercer, N. (1987). Common knowledge: The development of understanding in the classroom. London and NY: Routledge.

Elstgeest,, J. (1985). The right question at the right time. In Harlen, W. (Ed.), Primary science: Taking the plunge (pp. 36-57). London: Heinemann Educational Books.

Gardner, H. (1989). Zero-based arts education: An introduction to Arts Propel. Studies in Art Education, 30(2), 71-83.

Gardner, H. (in press). Assessment in context: The alternative to standardized testing. In B. Gifford (Ed.), Report of the Commission on Testing and Public Policy. Boston: Kluwer Academic Publishers.

Gee, J.P. (1989). What is literacy? (Technical Report No. 2). Newton, MA: The Literacies Institute.

Geertz, C. (1973). The interpretation of culture. New York: Basic Books.

Geertz, C. (1983). Local knowledge. New York: Basic Books.

Guthrie, G.P. (1985). A school divided: An ethnography of bilingual education in a Chinese community. Hillsdale: Lawrence Erlbaum Associates.

Hatano, G. (1981). Cognitive consequences of practice in culture specific procedural skills. The Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 4, 15-18.

Heath, S.B. (1983). Ways with words: Language, life, and work in communities and classrooms. New York: Cambridge University Press.

Hein, G. (Ed.) (1990). The assessment of hands-on elementary science programs. Grand Forks, ND: Center for Teaching and Learning, University of North Dakota.

Howard, K. (1990). Making the writing portfolio real. The Quarterly of the National Writing Project, 27, 4-8.

Inagaki, K., and Hatano, G. (1983). Collective scientific discovery by young children. The Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 5(1), 13-18.

Knorr-Cetina, K.D. & Mulkay, M. (Eds.). (1983). Science observed: Perspectives on the social study of science. London: Sage.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. American Education Research Journal, 27(2), 29-63.

Latour, B. (1987). Science in action. Cambridge: Harvard University Press.

Latour, B. & Woolgar, S. (1986) Laboratory life: The social construction of scientific facts. Princeton: Princeton University Press.

Lave, J. (1988). Cognition in practice. Cambridge: Cambridge University Press.

Lemke, J. (1990). Talking science: Language, learning and values. Norwood, NJ: Ablex Publishing Corporation.

Literacies Institute (1990). The Literacies Institute: Its mission, activities and perspective on literacy (Technical Report #1). Newton, MA: EDC.

Lynch, M. (1985). Art and artifact in laboratory science: A study of shop work and shop talk in a research laboratory. London: Routledge and Kegan Paul.

McKnight, C., et al. (1987). The underachieving curriculum: Assessing U.S. school mathematics from an international perspective. Champaign, IL: Stipes.

Medawar, P. (1987). Pluto's republic. Oxford: Oxford University Press.

Mehan, H. (in press). The school's work of sorting out students. In D. Boden & D. Zimmerman, (Eds.), Talk and social structure. Cambridge: Polity Press.

Michaels, S. (1981). 'Sharing Time': Children's narrative styles and differential access to literacy. Language in Society, 10(4), 423-442.

Michaels, S. & Bruce, B.C. (1989). Discourses on the seasons. (Technical Report). Champaign, IL: University of Illinois, Reading Research and Education Center

Michaels, S. and O'Connor, M.C. (1991). Literacy as reasoning within multiple discourses: Implications for policy and educational reform. Paper presented at the Council of Chief State School Officers 1990 Summer Institute.

Mohatt, G., & Erickson, F. (1980). Cultural differences in teaching styles in an Odawa school: A sociolinguistic approach. In H.T. Trueba, G.P. Guthrie, K. Au, (Eds.), Culture and the bilingual classroom: Studies in classroom ethnography (pp. 105-119). Rowley: MA Newbury House Publishers.

Moll, L. (in press). Bilingual classroom studies and community analysis. Educational Researcher.

Mullis, I. & Jenkins, L. (1988). The science report card. Princeton: National Assessment of Educational Progress.

Palincsar, A.M. and Brown, A, (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. Cognition and Instruction, 1, 117-175.

Philips, S. (1972). Participant structures and communicative competences: Warm springs children in community and classroom. In C. Cazden, D. Hymes & V. John (Eds.), Functions of language in the classroom (pp. 370-394). NY: Teachers College Press.

Resnick, L. (1989). Treating mathematics as an ill-structured discipline. In R. Charles & E. Silver (Eds.), The teaching and assessing of mathematical problem-solving (pp. 32-60). Reston, Va: NCTM.

Rosebery, A. and Warren, B. (1991). Video case studies of scientific sense-making: A constructivist approach to teacher development (Proposal). National Science Foundation.

Rosebery, A., Warren, B. & Conant, F. (1990) Making sense of science in language minority classrooms (Technical Report No. 7306). Cambridge, MA: Bolt, Beranek & Newman.

Rosebery, A., Warren, B., & Conant, F. (in press). Appropriating scientific discourse: Findings from language minority classrooms. The Journal of the Learning Sciences.

Schoenfeld, A. (in press-a). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), Handbook for Research on Mathematics Teaching and Learning. New York: MacMillan.

Schoenfeld, A. (in press-b). Reflections on doing and teaching mathematics. In A. Schoenfeld (Ed.), Mathematical thinking and problem solving.

Schön, D. (1983). The reflective practitioner. NY: Basic Books.

Schön, D. (Ed.) (1991). The reflective turn: Case studies in and on educational practice. NY: Teachers College Press.

Seidel, S. & Zessoules, R. (in press). Through the looking glass: The place of reflections in student learning. Paper presented at the Institute on New Modes of Assessment, Cambridge, MA.

Shymanksy, J. and Kyle, W. (1990). Establishing a research agenda: The critical issues of science curriculum reform. (Report of conference held April 8, 1990). Atlanta: The National Association for Research in Science Teaching.

Sizer, T. (1984). Horace's compromise. Boston: Houghton Mifflin.

Stock, P. (1990). Taking on testing: Chapter two. In G. Hein (Ed.), The assessment of hands-on elementary science programs (pp. 31-65). Grand Forks, ND: Center for Teaching and Learning, University of North Dakota.

U.S. Department of Education. (1991). America 2000: An education strategy. Washington, DC: Author.

Vygotsky, L.S. (1978). Mind in society. Cambridge: Harvard University Press.

Vygotsky, L.S. (1985). Thought and language. Cambridge: MIT Press.

Warren, B., Rosebery, A., & Conant, F. (1989). Cheche Konnen: Science and literacy in language minority classrooms (Technical Report No. 7305). Cambridge, MA: Bolt, Beranek & Newman.

Warren, B., Rosebery, A., & Conant, F. (in press). Discourse and social practice: Learning science in language minority classrooms. In D. Spener (Ed.), Theory and practice of biliteracy, Englewood Cliffs, NJ: Prentice Hall-Regents.

Warren, B., Rosebery, A., Conant, F, & Hudicourt-Barnes, J. (1991). Cheche Konnen: Collaborative scientific inquiry in language minority classrooms/A handbook for teachers. Arlington, VA: Development Associates, Inc.

Wells, G. (1981). Learning through interaction: The study of language development. Cambridge: Cambridge University Press.

Wolf, D. (1989). Portfolio assessment: Sampling student work. Educational Leadership, 46(7), 35-39.

Wolf, D. (1990, Spring). Toward a theory of creativity. Newsletter for the Society for Research in Child Development, 1-2.

Wolf, D., Bixby, J., Glenn, J. and Gardner, H. (in press). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), Review of Research in Education.

Response to Beth Warren and Ann Rosebery's Presentation

Ron Rohac
San Bernardino City
Unified School District, California

I come to you from about 15 years of the classroom wars. I am still in a classroom, and I have no scars or wounds to show for it at this moment because I've only been in class one day. The school district was nice enough to release me to come to this, and I am very honored to be here.

First of all, I thought I should give you some background as to my experience with limited English proficient students and then go from there in terms of discussing what implications there are with the paper, some of the problems she illustrated and pointed out, and then some methods of assessment that I use in my classroom so that my students will be able to tell or to show me that they understand the science content that I deliver to them, and their application of such science and information.

I have up to as many as 15 different languages in my classroom. I usually have a group of 25 students, at least that's set down by the district, which usually bulges to about 35, and by the time I get back on Monday, I'm sure each class will be about 50. Hopefully by that time, they will have divided the classes in half again and that I will have 25 students. They come to me from a range of backgrounds. Obviously, with different language backgrounds in different parts of the world, they generally come to me with limited English ability and I should tell you what that meant -- at least in my district -- when they asked me to do this kind of program. Like most new teachers to a program I was coerced into starting it. They told me that limited English meant that they did not speak English very well.

I grew up in Canada and taught in Canada for seven years. But when I moved to southern California, I found the students that I faced in a high school setting didn't speak English very well either, so I didn't see the difference at that point. Then they told me that the primary language was not English. So I understood that was the case and that shouldn't have been a problem. So I asked, "Exactly how am I to teach these students." They told me to speak slower, speak louder, and do lots of things. That's sheltered English. After a little bit more training and a lot more experience, we've come to revise that program substantially.

For the San Bernardino City Unified School District, a sheltered program is a content area where the youngsters are taught at grade level and English is not their primary language. I think it's really important that I emphasize that point again, that students are taught at grade level. There is no remediation of the course content whatsoever. My particular courses of study are physical science which are elements of chemistry and physics, and life science which is biology. The students are receiving ninth and tenth grade science credit, and they are working and functioning at that level and are achieving well beyond what most people expected them to do.

The paper, as I see it, raises two important problems for ESL students who are limited--and I don't like the words limited English students to be perfectly honest. The first word, limited, is not something that I appreciate. The first one was that science classes have not traditionally done what they're intended to do. I agree with that entirely, that science classes as they are traditionally taught are the drill and kill effect. That is, you will memorize a bunch of words and you will spit the words back to the teacher and if you do that you get an A, if you do a little bit less you get a B, and so on down the way. That's not the purpose of science education.

Second, for many LEP students, science teachers are not teaching science classes. That's a major issue and concern. It's usually left up to the ESL teacher--assuming science is in the curriculum. An ESL teacher is not a science specialist and my heart goes out to those people that are teaching those classes because they do not have the background experience or knowledge to truly teach science as it is designed to do. In the new framework which Dr. Warren has set out, it would not be possible for a non-science teacher to teach that class effectively. You would in essence be doomed to failure. The purpose of a science class should be to develop a way of solving questions. My particular idea behind teaching science is to instill the question, why, to my students.

Students should leave with an understanding of how to solve a problem but, most importantly, to ask a question why, and then go about their business of solving that particular question, or series of questions, to come up with answers that I'm going to pose to them. I agree with the idea that students should emphasize their particular opinions and their interests, but I am also a great believer in the formation of a curriculum that they must follow and, with proper teaching techniques, styles, and methodologies, the teachers can direct their students through elements of science and extend this particular methodology to the students' interest.

About 16 years ago when I graduated from college there was a wonderful new element of science education that was being purported and that was called the discovery method. A few years ago I read another article that talked about into, through, and beyond. I'm looking at another one called sense-making problems, or problem solving as sense-making ideas. And my question to the people involved in those things a few years ago or 16 years ago is this: If I have a 20-chapter curriculum to follow and I use only the discovery method to teach, then my youngsters will get to the end of Chapter One. They will not achieve the curriculum, and they will not be able to formulate the ideas and things set out by me, the district, and the state. I think there are some important issues to deal with, and science must be attested on these different levels.

Concerning the first business with the teacher, I wanted to talk a little bit about what ESL teachers are doing. I saw it as being a compliance issue or, should I say, an out of compliance issue? If ESL teachers are teaching science, then this particular district is not in compliance with the state recommendations. Science teachers are supposed to teach those courses. If I am to teach a math class, then the district will most likely slap my hands and get me back to science courses or social studies. Therefore, why are we expecting ESL teachers to teach a class that is very complicated and complex and definitely has all sorts of wonderful ramifications and they're just trying to struggle with the language. That's a whole game to themselves.

But the purpose behind what I'm supposed to do here today is talk about assessment and so I have some other things to deal with. A sheltered teacher will be given specialized techniques so that students can achieve the content. That is, they will provide comprehensible input. Listening to other things this morning, we seemed to have gotten away from the idea of something called, BIC's and CALP. Maybe that's very small potatoes in terms of this particular symposium but, as I understood it, my youngsters came to me with a basic understanding of English, or very minimal understanding, and the things that I'm going to teach them, the cognitive things, are the things that they're supposed to comprehend. my focus is basically on that particular level.

The techniques that I work with to gain this comprehensible input can be summarize into four major points:

One, we use things to visualize concepts, picture files, whatever it takes in essence to make an abstract concept concrete. That's what I'm really most interested in.

The second business of teaching is the development of hands-on activities and materials so that youngsters can go beyond what they see and understand and extend that particular concept to the application level. I am very concerned about Bloom's taxonomy in that youngsters will get into applications synthesis and evaluative components based upon things they have to do and construct. Cooperative learning, to take advantage of students' strengths and build on their weaknesses and to take advantage of the diverse backgrounds with which my students come to me. If we have 15 or 16 different languages in the classroom, we literally have a world of experiences, and their perspectives are different. And, if given the current situation, these different perspectives can be powerful tools for science education.

Then something else called guarded vocabulary, the method by which the teacher speaks, our rate of speech, our ability to enunciate words, to avoid idioms and colloquialisms, to use things in context efficiently, will allow my students to gain something called comprehensible input. In essence, my students will understand what it is I'm trying to present to them, they will be able to use that information and prove to me that they understand the concepts presented. The techniques described present a pragmatic methodology to teaching. Their goals are similar to the goals of the directions to developing scientific literacy. When I think of what goes on in the state of California now with the new science frameworks and the sheltered techniques that we use in the classroom, there are mirror images: one, to make content meaningful; two, to emphasize concepts rather than teaching fragmented bits and pieces of science; three, to develop and utilize skills taught to develop a creative and critical thinking level; and four, to teach vocabulary as needed to function in and around the concepts.

I will not teach words just for the sake of words; they have to have meaning behind them. As can be seen, the sheltered classroom focuses primarily in content. It is because of this focus that language can be acquired because language will have meaning. That's a key ingredient for me -- language will have meaning. It has been my experience that students develop science concepts and English without compromising the content.

Techniques used to assess students should reflect a teaching style used by the teacher. In this case, we must look for pragmatic ways to assess performance. Authentic assessment techniques allow students to demonstrate their knowledge. I have nine listed here and I would just like to go through them briefly.

One technique that I use is open-ended questions and open-ended activities. In open-ended questions, what I'm really most concerned about is that the students are going to tell me how they think they are processing their learning. That sounds like a lot of words but that's really the case. In open-ended activities, the student will demonstrate application. A nice example of an open-ended activity for my students is to hand my pairs of students pieces of aluminum foil. Their job is to tell me how thick it is. They have been worked through the areas of metric system in measurement, they have been worked through the areas of density, and they have all of these wonderful tools available because they've gone through that process and they've done all the measurement things. But it isn't enough. They have to be able to apply that information. So when I hand youngsters a piece of aluminum foil and ask them how thick it is, then they must be able to use that information. There are several possibilities that could be the correct answer depending on how the youngster thinks, in essence, his perspective and background, then it will be his solution. None of them can be wrong.

Another technique is the use of performance based tests which represent nearly 50 percent of my grading scale. Here, the youngster will show me what he/she knows, designing a human face based upon genetics information, building all sorts of different structures such as designing a cell, the components of a cell, are good examples of performance based activities.

So science can be tied to other curricula -- social studies, reading, writing, all are important. So they do not see science as being something else, we can't do this in class today because that's math and this is science class. What I usually tell my students at that point is, well, we shouldn't open the textbook today because we would be reading and that's English class.

Enhance multiple-choice questions. Here, an enhanced multiple-choice question represents the only kind of multiple-choice questions my students will see. Those particular questions, as such, being enhanced, use some form of the visual that is completely tied to the question. In other words, the question could not be answered without the presence of visual forms. Multiple-choice questions for the most part for my students are multiple guess. I am not testing their ability to read English, I am testing their science ability. So I try to avoid those.

Another technique which has gained lots of popularity in all sorts of subjects is the use of student portfolios. But there are teacher components which we call the evaluative component and student components which are the effective components. Students are responsible for inputting information into their portfolios. After all, they are their portfolios.

In this regard, I want to mention the use of interactive journals to practice writing. In a non-threatening way, students are going to be encouraged to write. The process the students are actually formulating, and the answer, is to define what science is. That's one of their jobs while I am gone for three days. They have been assigned to groups and they're to come up with a definition of what science is and that includes what will be included in this course, what they expect. Many of my students have never been in a science class. So it is with interest that I go back Monday to find out what they have written down.

Cooperative projects put kids in groups that will take them beyond their individual capabilities. And, by carefully designing those cooperative groups, students have become functional on a number of levels. But it is an exciting process to watch them go beyond the content as I expect to see it.

Finally, I recommend the use of anecdotal notes; things that I write down in class about "student talk." Dr. Warren had talked about student talk as being an important issue, and it is very important because: one, it develops concepts cooperatively; two, students think through problems; three, students express concerns and opinions; and four, students develop language skills. But there's a problem. In my particular classroom, English is not necessarily the language that the students discuss their work in; that's a major issue. If this is truly going to be a sheltered classroom, then the youngsters can function in whatever language suits them the best. As a teacher, I must be comfortable with the fact that they are working. It's been my experience that when students laugh and giggle in my physics class, I know it's not physics.

Student talk is important to the development of concepts, but a question to consider: Is language of the discussion important? I think not. I want my youngsters to struggle with the concepts of science; I do not want them to struggle with the concepts of English. So when they work in Vietnamese or Chinese or Spanish or Hungarian or whatever language, I face that particular day, or that year, it is of no interest to me. The students are functioning and working at their appropriate levels, and they go well beyond what they are capable of in English.

My conclusion, science or any other content-based class can be a powerful language acquisition device for potentially English proficient students. At the same time, it provides an opportunity for students to continue their education at grade level provided teachers do not remediate their courses but rather restructure their approach to teaching and assessment. Secondly, teachers, counselors, and administrators must remove the mind-set of remediating students listed as LEP. Finally, content-based classes should be taught by content-area educators and not ESL teachers. If these criteria are met, there are no limits for limited English students.

Response to Beth Warren and Ann Rosebery's Presentation

Sau-Lim Tsang
Northern California
Multifunctional Resource Center, Oakland

I am going to use a common sense approach in my short talk here.

First of all, I want to comment that the lecture format of this presentation is non-sense making. I am very interested in the "sense-making approach" (I am not sure "approach" is the correct word to use) described by Dr. Warren because it addresses one of the most important objectives of science education. That is, we want our children to be creative, to be critical, to be curious about nature, and to conduct scientific inquiries.

I'm especially interested in the title "sense-making." Actually, when I first heard this title last week, I asked a colleague whether he had ever heard of the term before. He said, "Yes, this is the latest thing, everyone is talking about it." I'm interested in it because science, when defined generally, is the understanding of nature. If you review the history of science, you find that the understanding of nature has always been guided by our perception and our sense.

In the early days, we made observations of the sky and we deduced that the appearance of the comet would be followed by an earthquake. In the Chinese folklore, the appearance of a bright star in the sky meant that a saint or an important person would be born. We drew relationship and conclusions by observing nature closely.

As our perception expanded (for example, when we invented the telescope), we were able to understand more natural phenomena. For example, we began to understand that the earth is revolving around the sun instead of the other way around. And when we invented the microscope and expanded our perception of small things, we also gained more understanding of the working of microscopic matters. Thus, the study of nature is guided by our perceptions, and sense-making is a very important part of science education.

I do have several questions about this sense-making approach. First, I am not clear about the difference between this and "scientific inquiry"; I do not have a clear definition from the paper. How is sense-making different from discovery learning and other similar teaching/learning methods? What happened to all of the science curriculum we developed during the 1960s, a period known as the golden age of science education, when the Congress provided large amounts of money for science curriculum reform and teacher training? One emphasis of the curriculum reform efforts of the 1960s was on discovery learning. Is there any relationship between this sense-making approach and the discovery learning emphasis of the 60s? Maybe Dr. Warren can address this concern in her paper.

Second, I want to know if there are any evaluations being conducted on the sense-making approach. Do we know if this approach is better than other approaches? We should know more about its effectiveness before the practice is disseminated.

Third, I am often confused by descriptions of innovative programs because they are often conducted by excellent teachers. I've been hearing a lot of descriptions about good practices based on one or two teachers. Are we talking about good practices or good teachers? Or is it a tautology? If I select a good teacher somewhere and I put a label on the approach she/he uses, does the approach then become an exemplary practice instantly based on its success with the teacher? I don't know why Jaime Escalante hasn't marketed his teaching method yet, since everyone knows how successful he is with his studies.

Fourth, I also want to know how the sense-making approach relates to children's stages of cognitive development, such as those proposed by Piaget to learning taxonomies, such as the one proffered by Bloom, and by extension, to the objectives for science education at different grade levels. In California, the State Curriculum Framework has developed a set of objectives for science education divided by grade level. For example, from K to third grade, the objective is to help students observe, communicate, compare, and organize objectives in nature; from third to sixth grade, students should understand the interaction and interdependence of systems of objects; from sixth to nine grade, they should explain phenomena through perceived changes in objects; then, from ninth to twelfth, they should use information to obtain further knowledge. How does the sense-making approach relate to these different objectives?

Fifth, I would like to find out how the sense-making approach facilities the ability of LEP students to overcome the language barrier. The paper provided descriptions of the language difficulties encountered by the LEP students and how these difficulties affected their access to the science content. But there was no discussion on how the sense-making approach helps alleviate these problems. For example, the author gave a description of a class project in which the students conducted a tasting test of water samples from different parts of the school building. Were the students who were less fluent in English left out of this project? Did they engage in discussion just as much as the others? I want to see the relationship between the sense-making approach and the education of language minority students.

The last questions I have is whether the sense-making approach can be disseminated to other teachers. If the approach is indeed successful, how are we going to disseminate this method? As was mentioned by the previous discussant, the sense-making approach is highly dependent on the teacher's scientific literacy and knowledge. A recent survey I read said that over 95 percent of teachers today are completely dependent on the science textbook to teach. They do not diverge from the textbook because they have very limited scientific knowledge. How can a teacher with limited scientific literacy adapt the sense-making approach?

I can give you an example. I once observed a sheltered English teacher giving a junior high school science lesson. The teacher was known to be an excellent instructor and well versed in the sheltered instruction approach. She was teaching a lesson on the effect of heat on matter. She was following the textbook and discussing the working principle of the thermometer-that mercury expanded as the temperature increased, thus raising the mercury column in the thermometer. Then a student asked a question: "Oh, yeah, we have a pot at home and the lid is always stuck. We can't open it. But if I put it in the oven, when it heats up, I can open the lid easily." The teacher said: "Yes, that is expansion." But another student asked: "The lid expanded but the pot also expanded. How come it is easier when both are expanded?" This was an excellent question which could be used as a lead-in to many hypotheses, experiments, and scientific concepts. However, the teacher ignored the question (probably because she did not have the scientific knowledge to respond to the question) and went on with the text.

These are all the comments I have regarding Dr. Warren's paper. For the remainder of the time, I am going to put forward some of my thoughts on the current "crisis" in science education.

Actually, this is the second crisis. We had the first crisis in 1957 when the Soviet Union launched Sputnik. We felt that we were losing the battle to the Russians and the federal government implemented a massive effort to improve math and science education. Numerous teacher training programs, curriculum development projects, and research projects were initiated and supported for over a decade. There were also many evaluations conducted with the curricula developed during this era. In general, the evaluations were based on these curricula. Students did as well as students in traditional curricula in factual learning and better in comprehension and concept application. I guess the culmination of all these activities was the moon landing in 1968. However, I am not sure whether putting our first man on the moon had much to do with the massive science education improvement program of the 60s. Instead, the fast advance of scientific research and development capability of this period may well have been the result of the large number of foreign scientists coming to our country.

The current crisis in science education, however, is quite different from the previous one. We are talking about how our industry is losing its competitiveness to other countries, about the fact that we need to modernize our industry and that our work force is not adequately literate in math and science to meet the needs of the changing industry. The last time, we wanted more scientists; this time we are talking about the general public, the general work force. We want them to be more scientifically literate. To ensure that our future work force possesses the required math and science literacy, we need to improve our math and science education.

This line of reasoning, though plausible, might not hold up when we compare it to schooling in Japan, purportedly our most fearsome competitor. The math and science curricula in Japan resemble what we had 40 years ago in the United States. Their work force is perfectly fine for their industry. Why do we have to change our math and science curriculum? That's something we have to think about.

I also want to give an anecdote about science in general. Last year I visited the Lawrence Berkeley Lab (LBL), where they have a special summer program to encourage young adults to enter science careers. In the summer program, they brought together the cream-of-the-crop students who showed interest in math and science related careers from all over the country, to introduce them to many exciting scientific projects that scientists were conducting at LBL.

I observed three young women working on a molecular experiment. One white, one an Indian from India, and one Hispanic. They all showed great enthusiasm and worked diligently at the experiment. Afterward, I asked each of them if the summer program experience helped them to select a science-related profession. The student from India said yes, she enjoyed science. The Hispanic woman said yes, she wanted to become a scientist. By the way, these two women are immigrants to the United States. The white woman, on the other hand, said that the summer program helped her to decide that science was not for her; she thought she would rather be a lawyer because science was just too tedious and boring.

This visit had me thinking about the lack of role models for our youths. When you look at TV today, there's not a single role model who is a scientist. Today's youths are greatly influenced by the mass media. When they look at TV, they see role models of lawyers, policemen/women, and some doctors (e.g. the Cosby show). But there are no scientists (though we do have Dr. Spock in Star Trek, who is not human.) How can we encourage our young people to be scientists?

I want to end with two examples. First, I did a research study six years ago with a group of high school students in San Francisco who are Chinese immigrants. One student especially impressed me. He came to the United States two years earlier from China and was a tenth grader enrolled in an Algebra II class. I asked him why he was taking the advanced math series. He told me that he really wanted to be a writer and his love was literature. However, his counselor told him that he had no chance in this country to be a writer and that he should study math and science to ensure a job in the future.

The second example is about one of my colleagues, who is here at this symposium, a Hispanic woman who grew up in the barrio. She told me that she grew up wanting to be a medical doctor. However, throughout high school, she was placed in a vocational track because she was told that she was not college material. Of course she was not able to study medicine. My colleague ended up a Ph.D from Stanford and she is one of the most capable people I know.

The first example may answer a question I often encounter-that is, why are there so many Asian-Americans in math and science-related professions? Even more disturbing, I am still unsure how I would advise this student if I were his counselor.

The second example illustrates the low expectations school staff hold for Hispanic students. The incident happened in the 1960s, but my current experience with schools suggests that a large segment of our school personnel still has very low expectations for certain groups of our children, and these expectations are often based on generalization without consideration of an individual student's background, ability, and potential.

Holistic Writing Assessment of LEP Students

Liz Hamp-Lyons
University of Colorado, Denver

Introduction

I thought I was here to campaign for the death of standardized testing, but it turns out that I'm here to say "I told you so." -- not to my physically-present audience, for I am among the converted, but to federal and state bureaucrats who have been antagonistic to or simply afraid of alternatives to standardized testing in general and to direct writing assessment in particular. I only hope that some of those people will read this book, and that this and the many excellent papers from the Symposium will not stay among the converted.

The irony of alternative assessment is that such a term should be needed. We have come full circle to the assessments of the turn of the century, writing prime among them. Is there a connection between the US's role as the multiple choice test capital of the world and an increasing anxiety abut declining educational standards? I think so. Is there a connection between declining literacy and the rise in social ills? I think there is. President Bush's little booklet, AMERICA 2000: An Education Strategy says:

- For too many of our children, the family that should be their protector, advocate and moral anchor is itself in a state of deterioration.
- For too many of our children, such a family never existed.
- For too many of our children, the neighborhood is a place of menace, the street a place of violence.
- Too many of our children start school unready to meet the challenges of learning.
- Too many of our children arrive at school hungry, unwashed and frightened.
- And other modern plagues touch our children: drug use and alcohol abuse, random violence, adolescent pregnancy, AIDS and the rest.

But few of these problems are amenable to solution by government alone, and none by schools alone. Schools are not and cannot be parents, police, hospitals, welfare agencies or drug treatment centers. They cannot replace the missing elements in communities and families. Schools can contribute to the easing of these conditions. They can sometimes house additional services. They can welcome tutors, mentors and caring adults. But they cannot do it alone. (p.10-11)

But this, it seems to me, is missing the point. Of course schools can't do these things alone; but neither can they achieve the AMERICA 2000 goal of universal literacy alone. Each requires the commitment of federal dollars. But AMERICA 2000 misses the point by a wide margin: It lays the blame for social ills at the doors of families and communities as though there were no record of the sociopolitical changes that have been primarily responsible for the increasing unemployment, poverty, exclusion and alienation lying behind these social ills. It blames "adult misbehavior" without acknowledging that not all the adults who've been misbehaving are in the children's homes or communities -- some of them are in high office, possessing the strings to the purses that contain the children's future opportunities. It lays the blame on the symptoms and not on the disease. And AMERICA 2000 goes on to propose curing the symptoms without attending to the disease.

AMERICA 2000 proposes that universal literacy is a more achievable goal than a nurturing family, a safe neighborhood and enough to eat. Happily, most of us will still be around in the year 2000 to assess the predictive validity of this proposal. My paper, then, is offered not as a claim that reformed practices in the assessment of writing will achieve the goals of AMERICA 2000, but as a range of options for improving writing evaluation as one very small practical contribution to one small part of the problem, within what I hope the National Education Goals Panel will swiftly realize must be a wholistic approach to problem-identification and solution-delivery to "make this land all it should be." (AMERICA 2000, cover page)

Holistic Writing Assessment

Definition "Holistic" writing assessment is the term used for tests which test writing wholly through the production of writing. While holistic writing assessments vary from national assessments such as the National Assessment of Educational Progress (NAEP) to teacher-made tests applied within a school building or even just one classroom, and from elementary school through college and graduate education, they all have certain things in common. A holistic writing assessment has at least the following five characteristics: First, each individual taking the assessment must actually, physically write at least one piece of continuous text of 100 words or longer and may write several pieces and/or considerably longer pieces. Second, while the writer is provided with a set of instructions and a text, picture, or other "prompt" material, she or he is given considerable room within which to create a response to the prompt. Third, every text is read by at least one, usually two or more, human reader-judges who have been through training for the scoring of writing in that context. Fourth, the judgments made by readers are tied in some way, tightly or loosely, to some common yardstick, such as a set of sample essays, a description of expected performance at certain levels, or one or several rating scales. Fifth, the readers' responses to the writing are expressed as a number or numbers of some kind, instead of or in addition to written or verbal comments; scores on the test are recorded and can be retrieved for review by higher or external authority as needed. It should be clear from the above that a writing test is a **performance** test.

Contrasts "Objective" tests are tests in which discrete elements such as the ability to recognize correct English word order, sentence structure rules such as tense maintenance, and vocabulary items dominate. Objective tests call on recognition skills not production skills: test takers select from a narrow set of choices created by the testers. While these skills may be related to proficient writing, as statistical studies have shown, most of us do not accept that they can represent what proficient writers do. The second kind, "analytic" tests require the test taker to write continuous prose, but instead of evaluating the text they use various count measures, such as mean number of words, word length, sentence length, number of errors per sentence, t-unit length, proportion of simple to complex structures, etc., which are claimed to be highly correlated with writing quality. Analytic assessment of writing does not involve the application of discourse-level measures of writing quality. As with objective tests, an increasingly large number of people, including teachers and researchers, do not accept that analytic measures can represent writing ability. The people who argue FOR holistic writing assessment ground their arguments in construct validity. They believe writing **must** be assessed with a performance sample.

Why assess writing with a performance sample? We live in a society that makes greater demands on the competencies of its members than at any time since the Industrial Revolution, and yet makes it easier than ever before for these members to exist at the fringes of that society in ways that are minimally functional, functional only because of the accommodation of the society to ever lower levels of functioning. I live in a city where more than half the Hispanic population do not complete high school, where 29 percent of the population as a whole and 9 percent of the college population are black. No longer, it seems, does the definition of a civilized society include education for all. What has this to do with writing assessment? Everything.

I am convinced that the methods of testing that have been prevalent in the last half-century bear some responsibility both for the declining educational and literacy standards in this country, and for the changing attitudes to education. "Education" has been reduced to that which can be tested in multiple-choice format, and which can be compressed into an item answerable in 60 seconds or less (since standardized tests depend in large measure on the number of items for their reliability). Teachers find themselves test-driven away from significant educational goals and toward limited sets of assessable knowledge. Children find themselves repeating similar problems again and again, in modes containing extremely low intrinsic motivation, because these are the forms used and areas covered on the test. "Education" no longer means the drawing out of talents, interests and capacities that its Latin origin suggests. An education no longer implies preparation for life and citizenship, for social and moral responsibility. Take a field visit to the pond, to carry out an experiment on specific gravity, or to observe the mating rituals of the crested grebe? Stop and write a poem about the clarity, the smells, the sounds of the day? Freewrite about the scariness of having a plane crash just blocks away from school? Learn to mix clay, to shape and bake it, to feel the simple beauty of it under your fingers, the satisfaction of making? Listen to stories of the lives of your grandparents, your neighbors? Read stories of the ordinary people who inhabit the land, who have made it what it is, the Polish, Greek and Asian early immigrants , the more recent Russian and Vietnamese immigrants, the Native Americans, the descendants of slaves, the Chicanos and Chicanas? Go out into the community and confront social issues, consider resolutions and begin action? Why? It won't be on the test. In my city, where the school-age population is more than half Hispanic American, Cinco de Mayo passed in my son's school with no celebration, no mention. His entire first grade year passed without a field trip.

There are two arguments levelled against holistic writing assessment, or performance testing of any kind. They are, that it is too expensive, and that the results are unreliable. In terms of expense, writing tests are not that much more expensive than standardized tests, since their higher cost for scoring is counterbalanced by the higher development cost for standardized tests. The development and use of writing tests also requires the involvement of skilled people in values clarification, test design, and scoring, bringing benefits in teacher skill development that must also be laid against the cost of direct writing assessment. Writing tests **are** more expensive, and they **do** demand the involvement of a large number of skilled people. But the evidence suggests to me and many others that our views of the cost/benefit of different forms of testing must be redefined to encompass not only test design and administration costs but also human costs and the practical economic consequences of each lost productive citizen. Human costs are not merely figurative, they are real. Teachers have always known this, but its truth has only recently been understood by business and industry, and it is this new understanding by corporate interests that lies behind the AMERICA 2000 initiatives.

The second argument, of unreliability, has been a difficult one for proponents of direct writing assessment to counter, in part because reliability is poorly understood. People are used to standardized tests. Test taking, and judgments about answers, go on invisibly, and the judgment process is automated. Questions are rarely raised about what goes on behind the scenes, and it is easy to forget, with standardized tests, that they too are subjective. The items are developed and selected by human judges; they are answered by human beings whose experiences and judgments may be different from those of the test designers; the "correct" responses are decided by human judges, as are the distracting "incorrect" responses. Standardized tests too, then, are not objective, but the scoring method obscures that fact, and people feel confident that they can depend on the scores to be "accurate." Standardized tests are "sold" to us because they are reliable: But this reliability means only that, once someone has decided what the answer will be, a clerical system ensures that only that answer is credited, giving 100 percent scoring reliability. No writing test can compete with that. And yet, scoring reliability is only one side of the issue. A test must not only test something consistently; it must also test the right thing. In this respect standardized tests are more difficult to pin down than performance tests are. Standardized tests claim to test large collections of skills with names like "language proficiency," which in fact has yet to be satisfactorily defined, or smaller sets of skills such as "grammatical competence," but can test it only by sampling a very small subset of the elements that together make up a language user's range of grammatical knowledge. Because they test only a very small subset of the possible microcomponents that make up any one of these larger skill/ability sets, the possibility of a "miss," of testing an element not known by this particular test taker, or of a "false hit," of testing an element this test taker is more familiar with than most others, is quite large. These decisions about test content are made by a small number of test designers, and they are made with a mix of expert judgment and individual variation that is much like decisions made by readers of writing samples. In fact, training for essay readers is highly developed and frequently written about and researched; the same is not true of training for item writers. But because on standardized tests the human judgment processes occur before the individual takes the test and not after, it seems less responsible for the individual results. This is clearly not true.

Educational testers call what testing does to teaching, good or bad, "washback" or "backwash," and it is true there are few empirical studies of it. But look at this country, and you see a giant laboratory, where the Method has been to construct an educational values system around standardized tests; where the Subjects have been America's school-age population; and the Results are before our eyes daily, on the streets and in the newspapers. Crime; drug abuse and drug pushing; teen pregnancy; gang violence; child abuse; spouse battering and family abandonment; homelessness; poverty. The highest neonatal mortality rate of any First World country. School dropout rates and illiteracy. College dropout rates and unemployment. Can we lay all this at the door of standardized testing? No, of course not. There are other well-documented sociopolitical factors which are in large part responsible. But I submit to you that the decreased attention to literacy in our schools, triggered by the decreased value placed on literacy by our school bureaucracies as represented by their mandatory testing policies, has led directly to decreased literacy at school exit and has been one factor in the rising numbers of semi-functional members of society. And this is a tragedy, not only a criminal waste of human resources, but a deprivation of joy, of growth, of self-knowledge, of opportunities for families to learn and love together. This tragedy cannot be measured. It is not limited to LEP students: It is a rot that has spread right through our education system and so through the society. Last night I walked past the Baptist Church just two blocks from this elegant hotel, where at 11 p.m. were twenty to thirty women and children crowded huddled onto the steps and in knots on the sidewalk. At 6 a.m. today I walked past the Department of Justice and read the words above the door: "Justice is the Greatest Purpose of Men on this Earth" and where I saw five or six men sleeping huddled on the warm air gratings of the building's narrow gardens. I passed the National Archives where I read the legend "The Heritage of the Past is the Seed of all our Futures." And I thought -- yes, and we are living it.

What part can alternative assessment, and holistic writing assessment in particular, play in providing a seed of hope for a more just future for our LEP, our minority, our poor and indeed all our children's futures? I believe it can play a part both through the message it sends to teachers, parents, and learners about what the society values, and through the concrete effects it has in necessitating a kind of "teaching to the test" which is congruent with the needs of the society and the individual future citizen.

In my view then **any** writing test is better than a standardized test. Later in this paper I make the specific argument that there is a form of holistic writing assessment that is ideally suited to LEP contexts. But before I do that, I want to describe the common writing assessment options currently in use. It is convenient to think of five components of a writing test: the writer, the task, the scoring method, the readers, and score reporting. While there is much that could be said on the subjects of writers, tasks, and readers (see Hamp-Lyons, ed. 1991), in this paper I focus on the scoring method

808

and score reporting, because I consider them to be particularly critical in the design of appropriate writing assessments for LEP students and for the evaluation of LEP education programs.

Scoring Methods for Holistic Writing Assessment

There is some confusion about the terms used in writing assessment, particularly the term "holistic assessment," and I believe it will be fruitful to establish and maintain a clear distinction between the terms "holistic methods of writing assessment" and "holistic scoring." There are several reasons for this confusion: One has been the desire by those in writing assessment to contrast all methods of evaluating writing through the judgment of actual samples of student writing with the objective and analytic methods almost universally used at the end of the 1970s, and still all too common today. The second reason is undoubtedly that direct writing assessment is still a very young field and there few people whose primary research interest lies within it, so that growth is both slow and somewhat haphazard. Although writing was almost universally assessed holistically in the early decades of the century, before the psychometric revolution of the 1930s, it was more of a "cottage industry," with few publications existing in the area. Once standardized tests were developed by and for the large government agencies—especially the Army and the intelligence agencies—research into writing assessment almost disappeared for a generation, and only concern about declining literacy levels in this nation brought it back. But the main reason for the confusion over terms is the difficulty of making clear to non-experts what a writing test is. To many people a writing test is simply the collection of writing, any writing, from students and then the making of impressionistic judgments about the quality of the results. Because the phrase "holistic scoring" has become the best-known one associated with writing assessment, it is not surprising that holistic assessment of writing and holistic scoring have become synonymous in the minds of many teachers and administrators. Add to this the failure of the writing assessment specialists to agree on terminology (a consequence of the youth of the field, referred to above), and the problem is difficult to eradicate. The distinction between holistic scoring and holistic methods of writing assessment is an important one. In a classic paper, Charles Cooper (1977) defined holistic evaluation as:

> any procedure which stops short of enumerating linguistic, rhetorical, or informational features of a piece of writing. Some holistic procedures may specify a number of particular features and even require that each feature be scored separately, but the reader is never required to stop and count or tally incidents of the feature. (p. 4)

This is the definition of holistic assessment used in this paper. "Holistic scoring," "primary trait scoring" and "multiple trait scoring," are all holistic methods for making judgments about writing, as is portfolio assessment with which I close my exploration.

Holistic Scoring

Holistic scoring seems to have been established independently in two similar forms in Britain and the United States, by Wiseman and his colleagues in England and known at that time as the "Devon method" (Wiseman, 1949), and by Educational Testing Service in the United States, best known through the work of Godshalk, Swineford, and Coffman (1966). In holistic scoring (or rather, in focused holistic scoring, the usual method currently) written texts are collected from test takers, usually responding to a quite general question or "prompt" within a limited time frame of 30 to 50 minutes. These are submitted to readers for scoring; readers usually meet together for training and scoring, although in many local holistic scorings readers take essays away to score them. Training is generally fairly limited, typically a session of two to four hours, and generally proceeds by referring immediately to essays and the writing standards they illustrate. There is a scale of some kind, most often running from 1 to 6 (with 6 usually being high), "benchmark" essays are used to show what an essay at each score level looks like. Readers read practice essays and try to match the "expert" scores previously assigned to those essays. The theoretical foundation upon which holistic scoring rests is that readers make judgments of texts as a whole: that they are unable to separate out facets or parts of the essay and identify them. While proponents of holistic scoring argue that holistic scoring "reinforces the vision of reading and writing as intensely individual activities involving the full self" (White, 1985, p33) and that any other approach is "reductive," ultimately agreement on scoring standards is typically reached by each reader adjusting her scores to try to come closer in line with the other readers in the public context of training. Further, holistic scoring requires agreement between readers to be generated from trial scoring of sample papers, and thus depends on the readers involved on a particular day reaching an accommodation among them for the standards they will apply on that occasion. The weaknesses of this approach, both for equitable student evaluation and for program evaluation, are immediately obvious. Adaptations have arisen, most notably the development of essay scales and/or rating guides to accompany holistic scoring sessions, resulting in what is known as "modified holistic scoring" or "focused holistic scoring", and testing agencies, especially Educational Testing Service, have refined the technique into a very efficient and accessible tool. But holistic scoring still yields only one score to express the quality of the student's text.

Figure 1 is an example of an actual writing assessment question used in a statewide writing assessment at eighth grade level, and the scoring rubric, or guidelines, used to score student writing on the prompt.

Figure 1
Holistic Scoring

Task:

We are beginning to understand how important it is for everyone to help protect the environment. What can your school and your class be doing to help the environment?

Rubric:

none

Scoring Instrument:

6	High/Excellent
5	Good
4	High Average
3	Low Average
2	Weak
1	Low/Very Weak

Monitoring for reader reliability is facilitated by the use of two readers for each paper, and readers' scores are correlated. The kind of reporting on the performance of individual students that is possible is shown in Figure 2:

Figure 2
Score Reporting (1) Students

(Class X, Grade 8

COMPOSITION

Adams, J.J.	4
Brown, C.	3
Dong, K.K.	2
Gonzales, R.L.	1
Hunter, W.	5
Jackson, J.	1
Nguyen, M.	2
Rogers, B.	4
Smith, D.	4
Santiago, D.	3
Taylor, B.	3
Weissbaum, E.	5
(etc)	

There are a number of serious problems with holistic scoring in any context, but these problems are especially serious in ESL writing assessment contexts. Chief among these is that holistic scoring is not designed to offer correction, feedback, or diagnosis (Charney, 1984). The integration of evaluation and education is being increasingly recognized in all spheres, and the trend is certainly toward assessment instruments that can inform pedagogical decisions in quite specific ways: This is simply not possible with holistic scoring. We are increasingly coming to view this as a severely limiting feature of holistic scoring, and to demand a richer definition of a "valid" writing assessment. For LEP and other special educational needs students in particular, diagnostic feedback and correction have a central educational role to play. Many LEP students have had only limited exposure to instruction in English, and are only part way through their individual development of their potential mastery of English. Given appropriate instruction, interlingual development remains a real possibility for most of these learners. As Figure 2 suggests, a single score does not provide sufficient information for the student, the teacher or the administrator to decide on the best use of teaching provision in the form of course placement or curricular options, or to set up plans for special services such as tutoring, conferencing or workshops. These services can be especially helpful to LEP students.

Another weakness of holistic scoring is the limited potential it offers for meaningful program evaluation. Suppose two classes in neighboring schools each use the same holistic writing assessment: the hypothetical data in Figure 3 might result:

Figure 3
Score Reporting (2) Program

SCORE	CLASS X (N=30)	CLASS Y (N=30)
6	0	2
5	2	5
4	8	13
3	13	8
2	5	2
1	2	0
	etc	etc

The two classes at the same level have very different results: that much is clear. However, the holistic score data provide no clues as to **why** that might be. Without a more fully-fleshed picture, any generalizations about the effectiveness of curriculum, materials, or teachers would be foolhardy.

Primary trait scoring

A second kind of holistic writing assessment is primary trait scoring, which is in fact, despite its name, more than a scoring method. Primary trait scoring is based on a view that one can only judge whether a writing sample is good or not by reference to its exact context, and that appropriate scoring criteria should be developed for each prompt (Lloyd-Jones, 1977). Primary trait scoring responds to what we have discovered about the influence of task and purpose on any learner's writing, by paying close attention to task specification and to establishing close congruence between writing goals, task demands and scoring. The theory is that every type of writing task draws on different elements of the writer's set of skills, and that tasks can be designed to elicit specific skills. One task might, for example, be designed to elicit the ability to write a formal letter of complaint, and another might elicit persuasion. Primary trait scoring also emphasizes appropriate content, and each task would be expected to elicit certain specific content depending on the exact topic and wording of the prompt. The primary trait scoring guide consists of: (1) the task, (2) the statement of the primary rhetorical trait to be elicited, (3) an interpretation of the task hypothesizing writing performance to be expected, (4) an explanation of how the task and primary trait are related, (5) a scoring guide, (6) sample papers and (7) an explanation of scores on sample papers. Clearly, development of the scoring guide and development of the prompt go hand in hand. I am going to take as my example, here and in the next section on multiple trait scoring, the same example I used above, and sketch out for you how it might be developed into a better instrument using the primary trait approach or the multiple trait approach. I will not be able to offer you a full instrument because the development of a good writing assessment instrument is a skilled, careful, and time-consuming process, and one that depends absolutely on extreme responsiveness to context. These examples were constructed not for a real assessment but purely for the illustrative purposes of this paper. The examples I give should not, therefore, be taken as examples of excellence but as examples of the shape and direction that excellence might take. Consider Figure 4:

Figure 4
Primary Trait Scoring

Task:

We are beginning to understand how important it is for everyone to help protect the environment. Write a letter to your school principal making some suggestions about what the school and your class could be doing to help the environment.

Rubric:

When you are writing your letter remember that it doesn't help just to complain. You need to have some practical and well-described suggestions for how the school, and your class in particular, can take action to make a difference.

Trait Specifications:

PRIMARY TRAIT= suggesting a solution to a problem

TRAIT DESCRIPTION: The trait requires the identification of actual areas of present environmental concern that relate to the activities of a school (e.g., waste paper disposal). It requires specific language in identifying a problem area and in suggesting a solution (e.g. composting; paper recycle boxes in each classroom, and a class rota of recyclers). It requires use of clear structure to signal a suggestion, e.g., "I think we should..." "What we could do is...." It requires a clearly-made connection between the problem (e.g. a lot of paper gets wasted in schools) and the suggestion for a solution (e.g. recycle boxes), such as, "If we xxxxxx then yyyyyy would no longer happen" or "Using yyyyyy would mean that xxxxxx is not as bad as it is now."

Figure 4 (Continued)

Scoring Instrument:

6 High Writer identifies a real problem in school buildings and names it appropriately. She identifies a reasonable way of dealing with this problem. She shows how it would be possible for the class or the school to put the proposal into action with the resources already available, or she shows how it could be done with only minor additional resources.

5 Good (would be added)

4 HiAv (would be added)

3 LoAv (would be added)

2 Weak Real weaknesses are evident in identifying a problem and suggesting a solution. There is no attempt to show the proposal could be put into action.

1 Low (would be added)

Figure 4 shows, first, a revision of the task in Figure 1: the revision was necessary to fit the more specific tasks implied by the primary trait approach. Then, the trait is named and characterized. The scoring instrument has the same six levels as in the holistic scoring example, but this time a fairly detailed statement of the expectations on the trait to be assessed is provided (I have completed only two of the levels, for the purpose of illustration: note again that is not an operational instrument). When scores are reported for students and groups of students, still only a single number is reported, as shown in Figures 5 and 6, but the numbers are more meaningful than scores from a holistic scoring because they apply **only** to the skill or trait that was assessed. The opportunity to use the language of the scoring instrument to report individual student performance is an important benefit of primary trait scoring, especially in the LEP context. Parents of LEP children are usually LEP themselves, and anxious about their children's ability to succeed in school. Descriptive reporting permits them to see not only a number, interpretable only by reference to some "norm," which in mainstream classrooms is a native speaker "norm," but also some real explanation, which they can read or have a more fluent English speaker read for them, which reports their child's performance against a criterion, against expectations for real language use.

Figure 5
Score Reporting (1) Students

Either

Same as Holistic Scoring

Or

by text description, e.g.:

Farizah's score was 3: she has shown that she can identify a problem and name it but not describe it in full detail with clarity or suggest a reasonable solution to it.

For program evaluation primary trait scoring also offers the possibility of a more explanatory model, as Figure 6 suggests:

Figure 6
Score Reporting (2) Program

Either

same as Holistic Scoring

Or

by text description, e.g.:

In Class X most children identified a real environmental problem and suggested a solution. Five children suggested solutions that were not realistic. No child was able to show convincingly how the solution could be put into effect within the school's existing resources by providing full detail of the operation of their solution. The papers in the middle (levels 3 and 4) were characterized by vagueness of content, etcetera.

In Class Y, two children achieved the highest score by demonstrating a convincing and realistic implementation of the solution to the problem; several other children made a fair attempt at doing this but omitted some important aspect of a workable solution, etcetera.

I believe you can see that the primary trait approach permits a much richer picture of what children have done and how well than does a holistic scoring. The limit is that this information is available only for a single trait, but when students are given several primary trait tasks, the several scores that result can provide a rich diagnos-

816

tic picture of where that student's strengths and weaknesses lie, and this diagnostic information can be very useful to teachers and administrators as well as to the students themselves. Because of the careful development and detailed specification of the trait and the involvement of teachers and essay readers in test development, when readers use primary trait scoring, they make judgments with the support of an instrument that gives very clear and strong guidance, and the social pressure of the holistic scoring session can be avoided. But the advantages of this ecologically rich assessment are bought at the cost of an expensive development procedure. Whereas when most schools and colleges use a holistic scoring procedure, they transfer and adapt one from a large testing agency with expert personnel and a development budget, the principles of primary trait scoring make this impossible. The competencies specified and tested must be those found to be salient for the context in which the writing assessment takes place, which means very careful needs assessment must precede the test development. In the primary trait method, every writing task requires its own primary trait scoring guide. Not only must each school and college develop its own prompts and primary trait scoring guide, it must do so with almost the same expenditure of time and expertise for every new prompt.

As I developed writing assessment instruments, first for large scale second language writing contexts, then for a first language plus advanced ESL population, I looked for a compromise approach between the rich detail and uncompromising specificity of primary trait, which was beyond the financial possibilities, and the cheap but unacceptably uninformative holistic scoring approach. Building on the principles of primary trait scoring and rather outdated work in analytic scoring, and stimulated in particular by the work of Jacobs et al (1980), I developed what I have called a "multiple trait" approach.

Multiple Trait Scoring

The basic concepts of context-appropriate and task-appropriate criteria that underlie primary trait scoring underlie multiple trait scoring also, and I owe the concept of multiple trait scoring directly to Lloyd-Jones' primary trait approach. The development of multiple trait scoring procedures has been motivated by the desire, first, to find ways of assessing writing which in addition to being highly reliable would also provide some degree of diagnostic information, to students and to their teachers and/or advisers; and second, to find ways of assessing writing with the level of validity that primary trait scoring has, but with enough simplicity for teachers and small testing programs in schools and colleges to apply in the development of their own writing tests. While I have developed multiple trait instruments for English L1 contexts as well as for LEP contexts, and believe in their great value in both, I am convinced that limited English proficient students stand to benefit particularly from a multiple trait form of writing assessment.

"Multiple trait scoring" implies giving separate scores for more than one facet or trait on any single essay. When proponents of holistic scoring object to methods that do this, they are usually reacting against the "analytic" scoring used in the 1960s and 1970s, which focussed on relatively trivial features of text (grammar, spelling, handwriting) and which did indeed reduce writing to an activity apparently composed of countable units strung together, hence the label "analytic," which came to have a derogatory connotation in writing assessment.

But what I am calling multiple trait scoring procedures are very different from the old analytic scoring. Like primary trait scoring, the multiple trait procedure is an approach to the whole writing assessment and not only the scoring. Reader training is the norm in all writing assessments these days, but a multiple trait procedure goes beyond this to include reader involvement in instrument development as a vital components. Like primary trait instruments, multiple trait instruments are grounded in the context for which they are used, and are therefore developed on-site for a specific purpose with a specific group of writers, and with the involvement of the readers who will make judgments in the context. Each is also developed as a response to actual writing on a single, carefully specified, topic type. However, because multiple trait instruments, at least as I have designed them, unlike primary trait instruments do not contain any content specifications, multiple trait scoring instruments can be applied to a range of prompts, as long as those prompts fulfil the initial design criteria for prompts for which the multiple trait instrument was developed, and as long as the context remains essentially unchanged. This makes them more viable for small but committed groups of teachers to develop, pilot, and monitor in their own context, thereafter adding new prompts and paying close attention that new prompts pursue the same writing goals as the original prompts. Of course, multiple trait instruments can be developed that **do** include content specifications, but the amount of work in both development and in training for scoring would be very great. Increasingly, the trend is to develop multiple trait scoring instruments to fit a particular view or construct of what writing is in this context, and to reflect what it is important that writers should be able to do with the written language. "Ideas" are found to be a salient trait in most contexts, but this trait is generally judged in the general rather than the specific (that is, of the nature of "pertinent and convincing ideas," "plenty of relevant ideas," "adequate quality of ideas," etc., rather than "contains ideas a, b, c and d" or "contains ideas a and b but not c or d").

Each of the characteristics of multiple trait scoring I have made brief reference to above is, I think, a significant difference between holistic scoring and multiple trait assessment. The on-site, contextual development of prompts and trait descriptors cannot be illustrated in a paper, but Figure 7, which shows our task again, this time in a multiple trait context, does suggest some of the outcomes to be expected of that development process. Note the explanatory rubric that students receive accompanying the task. Note also the task specifications which guide not only the readers' movement toward shared expectations on this task, but also the processes of communal development of new prompts of the same task-type to be scored on the same scoring instrument.

Figure 7
Multiple Trait Scoring

Task:

We are beginning to understand how important it is for everyone to help protect the environment. What can your school and your class be doing to help the environment?

Rubric:

There are a lot of different ways schools can help the environment, but you will do well on this task if you think of one of them, explain it clearly and show clearly what action the school could take. Be specific and realistic in explaining how your proposal would work.

Task Specifications:

Problem—>Solution. These tasks require the writer to make a clear specification of a/the problem, putting it into the appropriate context. They also call for a textual connection between the problem and a proposed solution. The solution should be explained in enough detail to give it credibility, and it should be convincingly argued. Opposition to or minor flaws in the solution need not be addressed.

Figure 8 shows the beginnings of a multiple trait scoring instrument for scoring this prompt and task-type. Note that, as I have stressed above, development of a multiple trait instrument should be a communal process; certainly it is a time-consuming one. In pursuing my purpose of illustrating the differences among writing assessment methods I have taken a prompt from a holistic scoring and adapted it within each of the methods. Therefore I have only begun to sketch out how trait descriptions might look in the multiple trait approach. To do more would not only be too time-consuming for merely illustration purposes: it might also mislead readers to see this as an actual instrument that might be taken and used in a real assessment context. For a completed, piloted, and validated multiple trait instrument, I refer you to Appendix A and B.

Figure 8
Multiple Trait Scoring Instrument

	Trait 1	Trait 2	Trait 3	Trait 4
Score	Problem/Solution text structure	Reasonable content	Development of specifics	Control of the language
6	Problem stated before solution; suggestion made before explanation. Text elements are logically related throughout.	Both problem and solution are reasonable and significant.	Neither problem nor solution is vague. Each is clearly explained. The proposal for how the solution would work is clear, detailed and rational.	Any language problems are too minor for the reader to notice.
5				
4				
3				
2				
1				

There are many positive differences between multiple trait scoring and holistic scoring, but the most obvious difference, and probably the most important, especially in the LEP context, is that in multiple trait scoring more than a single score is generated and reported. In the Michigan Writing Assessment, for example, the instrument I developed generates four scores, all of which are used in decision making, and the descriptive correlates of three of these are reported to the student herself or himself as diagnostic feedback and as a textual explanation of placement in the writing program. (Appendix 1 and 2) Like primary trait scoring, multiple trait instruments focus only on the most salient criteria or traits for the context, and do not claim to assess every facet of writing competence that may appear in the student's writing. This means that careful test development is essential to establish what features are salient, and this development must focus on careful data collection in and about the writing situation where the test is located. At the eighth grade, for example, participant observation might reveal that teachers considered the ability to see problems outside the self as a salient feature, and one trait in a multiple trait instrument might attend to how far the writer builds comments about how individual choices lead to problems for larger groups into her text. Related to this is the important trait of problem solving, and another trait might focus on the ability to propose and describe solutions to problems. Another salient feature at this level is likely to be evidence of the student's developing control over sentence structure, the ability to use compound and complex sentences in appropriate rhetorical contexts. Discoveries about what features are salient may be made through discussions with teachers, practice scoring, and discussion of a range of essays, study of the marginal notations on in-class writing from the same context, discussion with teachers in other subjects in the school about the strengths and weaknesses **they** note in students' writing at that level, and so on. But the outcome of this data collection stage is always a statement of the salient features to be assessed in this context and on this occasion. The principles and the basic procedures do not change from the college context through the school grades because of its context-dependent nature, this approach is suitable for all levels and situations where writing is assessed.

Figure 9 attempts to illustrate the richness of information about individual performance that can be obtained from a multiple trait assessment (refer back to Figure 7 for the trait explanations):

Figure 9
Multiple Trait Score Reporting

(1) STUDENTS:
EITHER Numerical, e.g.:

Class X, Grade 8

	Problem/Solution	Content	Development	Language	TOTAL
Adams, J.J.	4	3	5	4	4
Brown, C.	2	3	3	3	3
Dong, K.K.	2	5	2	1	2
Gonzales, R.L.	1	1	2	1	1
Hunter, W.	5	5	3	6	5
Jackson, J.	1	1	1	1	1
Nguyen, M.	2	2	2	2	2
Rogers, B.	6	5	4	3	4
Smith, D.	4	4	4	5	4
Santiago, D.	3	5	3	2	3
Taylor, B.	3	3	4	2	3
Weissbaum, E.	5	6	5	4	5
(etc)					

Figure 9 (Continued)

OR by text description, e.g.:

> Bajni's writing showed excellent control of problem/solution structure, with clear textual relationships. Bajni offered a reasonable problem and solution, but one or both of them might have been more significant. Bajni developed the material fairly well, although there is room for more detail in the writing. Bajni's language control is still developing, and readers are aware of a number of problems of use of language in the writing.

To recap: A multiple trait instrument is an attempt to build up a scoring guide that permits readers to respond to the salient features of the writing whether these are all at the same quality level or are at several different quality levels. The essential characteristics of the multiple trait instrument are its grounding in actual reading data from the context where decisions are to be made; the selection of facets of writing quality in that context shown to be most salient by readers in the context, which in turn permit the reader to attend to what is salient on future reading occasions; and the provision of scores on each of these facets for use in decision making such as acceptance into a program or placement within a program, or in diagnosis of specific problems to be addressed within the instructional context.

Multiple Trait Scoring and LEP Writers

Writing assessment measures very like multiple trait assessment have been used for over a decade now in assessing the writing of second language English writers. Jacobs, Zinkgraf, Wormuth, Hartfiel and Hughey (1981) developed the "ESL Composition Profile," a scoring procedure containing several clearly articulated scales for the scoring of different facets of writing and introducing the term "profile" which I have found so useful. The ESL Composition Profile became deservedly very widely-known and emulated, and has been transferred into and is still used by many college-level ESL programs today. Jacobs et al., worked as a team, they conducted a detailed literature survey, and piloted their instrument carefully; they did not, however, collect observational data from which to build their instrument: rather, they began with criteria previously established for the test and expanded and refined them. Weir (1983) developed a writing test for postgraduates in Britain based on extensive questionnaire data from many British universities coupled with observational studies of faculty at the University of Reading. The collecting of empirical data and building of scales in response to it takes Weir's work closer to the development process I imply by the use of the term "multiple trait," but Weir did not work with readers as he developed his scoring procedure. Purves (1984) and a team of International Education Association researchers developed a large and complex set of scales for measuring the writing of high school writers in many countries against a common set of values. Although a number of useful insights have come from this work, the size and complexity of the instrument have meant that they are not used outside the IEA-funded studies. I have already referred to some of the insights which came from my work as a consultant to the British Council developing multiple trait instruments for two task types used in assessing the writing of ESL postgraduate entrants to British universities (Hamp-Lyons, 1984, revised 1986).

Each of the studies I have referred to has shown that reliable scores can be obtained using well-designed methods of holistic assessment that are more detailed than holistic scoring -- by which is meant a multiple trait scoring procedure with carefully developed and monitored prompts, a multiple reader system, reader involvement in the development process, and thorough initial and refresher reader training. Each of the studies I have referred to has focused on the assessment of the writing of nonnative writers of English.

Every writer would benefit from sensitive and detailed feedback on their writing, but LEP writers have a special need for scoring procedures that go beyond the mere provision of a single number score. First, for reasons that at present are unclear, LEP writers often acquire different components of written control at different rates. Every instructor of second language writers has encountered those students who have fluency without accuracy and those with accuracy but little fluency. We also sometimes see writers who have mastered a wide vocabulary but markedly less syntactic control; or who have syntactic control not matched by rhetorical control; and so on. With second language writers who already have some mastery of a specialized discipline, it is quite common to encounter texts that show very strong content while grammatical and textual competence lag far behind. De Jong & Henning (1990) have suggested, based on preliminary analysis of a very large data set, a pattern of language acquisition in which absolute non-users of the language have a single dimension to their performance -- zero on everything, and at the highest levels their performance on different tasks and skills once again converges so that they again show a single level of competence, this time a high one: But in between, they advance in different areas more quickly than in others (depending on language background, exposure to English, school and social context, and many other factors), so that their test scores appear divergent and multidimensional. We need writing assessment measures that provide the level of detail that allows such disparities to emerge.

Another argument for the use of multiple trait assessment is that the chances of significant improvement in writing, and the speed with which this can occur, are both greater for LEP writers than for most L1 writers. On one hand, growth in writing proceeds slowly for most first language writers of English after about eighth grade. Second language writers, on the other hand, are in the process of developing their language skills, of acquiring new areas of control and expanding their confidence in areas where they already have some control. LEP writing teachers have the joy of seeing their students make real progress, often in rather short periods of instruction, at any age. The potential for using writing assessment instruments to measure the real language gain of second language learners over a course of instruction (that is, achievement testing) is very real, but once again this means that a detailed scoring procedure is needed.

Another reason for a special kind of scoring of LEP writing is to help ensure that scores reflect the salient facets of writing in a balanced way. LEP writing typically contains significantly more language errors than L1 writing (McKenna and Carlisle, 1991), and the danger is that readers might respond negatively to the large number of grammatical errors found in many second language texts, and not reward the strength of ideas and experiences the writer discusses. This is especially likely to happen where LEP writers are part of a larger test candidate pool containing mainly L1 writers, and readers don't have special training in teaching LEP writing. The opposite can happen too: If the assessment emphasizes ideas and formal argument structures, readers may not attend sufficiently to language errors that would be seriously damaging in most school and college courses. Holistic scoring would obscure a pattern of consistent overemphasis or underemphasis on basic language control. These problems can be minimized by the use of a multiple trait instrument in which this facet is a trait to be judged, together with other facets found to be salient in the context, and where readers are freed to attend to the multidimensionality of ESL writing.

Advantages of Multiple Trait Assessment

While multiple trait instruments are less costly than primary trait instruments because they can be used with multiple prompts that fit the design parameters for the instrument, they are considerably more costly than holistic scoring because of the extensive development efforts involved. What, then, are their advantages?

Reliability When the scores on the multiple traits are combined to create a single composite score in use in making an administrative decision, that single score is highly reliable. In a study of an adapted version of the New Profile Scale developed for the British Council as applied to ESL essays from entirely different contexts, Grant Henning and I found that composite scores were consistently above

.90. (Hamp-Lyons & Henning, 1991). The use of composite scores increases reliability as follows: Assume a multiple trait scoring method with four traits: thus four scores are collected from each reader. Assume also that each essay is scored by two readers, as is the most common practice in writing assessment programs. The result is eight scores, four matched pairs. We may then obtain correlation coefficients for each pair of scores: each of these uncorrected correlation coefficients is an estimate of the reliability of the score on that trait if a single reader were to read each essay and give a score. Because two judges are used, scores will in fact be more reliable than that estimate, and we may use Spearman Brown's prophecy formula, also known as correction for attenuation, to estimate the increase in reliability[1]. Most programs also use a third reader in cases where the first two readers are far apart in their judgments; the way these third scores are used varies, but their result is an adjudicated score that is theoretically closer to a "true" score than the first two scores alone. Generalizability theory (Bachman, 1990) would fulfil the same function, but correction for attenuation can be done quickly by hand by the least statistically literate among us. Thus the multiple trait procedure possesses psychometric properties that enhance the reliability of single number scores built from its components, which can be used for making yes/no decisions such as whether or not to accept a candidate into a program of study where writing competence is required, and for setting cut points such as the level below which a student should be placed into a remedial writing program. While single scores are often used for these purposes, the reporting of the trait scores seems to me to be a vital part of the multiple trait assessment; I will discuss this in detail in the section on Increased Information below.

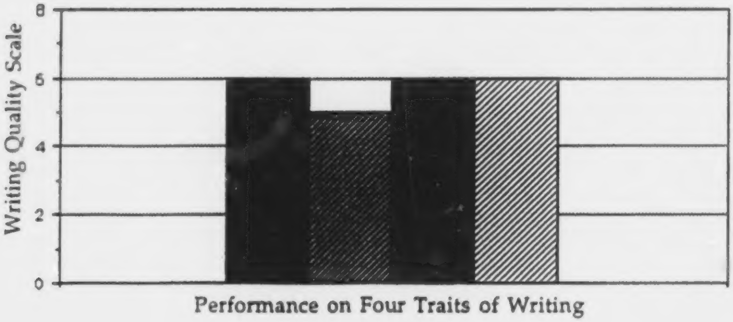
Validity No test can be valid without first being reliable: only when we have stable score data to look at can we usefully go on to ask questions about validity. But reliability does not imply validity: to judge validity, we need to look at other kinds of data. Following Anastasi, 1982, I take construct validity to be the overarching validity, and it is this type of validity which is central in writing assessment. When a test accurately measures the behavior which defines the construct, it has construct validity. Subsumed within this is content validity, for the traits in the multiple trait instrument derive from fairly concrete expectations in the college or workplace setting. Construct validity and content validity come from careful observation of a context and the shaping of the instrument to fit with those observations. If, when test design is complete, others can look at a test exemplar and see in it the appropriate behavior and values for the context, the test has achieved ecological validity. To ensure content and construct validity, test developers must pay careful attention to the evidence for what is valued in writing in the context to which the writing test applies, design prompts to elicit that kind of writing and scoring procedures to judge those values and ensure that readers keep those values in mind. These judgments of prompts and scoring procedures are in large part content validity judgments (note that content validity can really only be measured by expert judgments). Cronbach (1949:48) called this "logical validity." This must be coupled with a clear sense of what is involved in the construction of written discourse, of the limitations imposed by the assessment medium -- keeping in mind what it means to write in these circumstances. The text construction in a one-hour impromptu is, after all, a very different matter from the text construction that is possible in a take-home assignment from a course. To then show empirical validity involves statistical validation to discover whether scores are closely related to other measures which are already known to measure the same, part of, or closely related, skills or behavior. This statistical validation is rarely done outside large testing agencies which employ full-time statisticians and researchers, and I would refer you to the Research Reports of ETS for examples of empirical validation.

Increased information A key statistical question that must be resolved when using a multiple trait scoring procedure is whether scores should be combined and if so, how. If diagnostic information is part of the purpose of assessment, clearly, each of the trait scores should be reported separately. If reliability is key, trait scores when combined result in highly reliable scores. In combining scores, we do not know enough (and may never know enough) about how facets of writing weave together and in what proportions, so that decisions about combining and weighing scores are always based on presuppositions and prejudices. If score combining is essential, in my view the safest way to combine scores is to weight each facet equally. If a development team feels a strong urge to weight one facet more heavily than others, that may be an indication that for this context a focussed holistic scoring would be sufficient. Score weighting for purposes of obtaining a single score should always take place with the advice of a statistical expert.

But it is when multiple trait scoring is combined with profile reporting that its chief advantage becomes clear. Profile reporting is the reporting of all the separate trait scores rather than, or in some contexts in addition to, a composite score. Scores exist not simply to **assign** decisions but also to **communicate** decisions. Scores are information which can be shared with the writers, their academic advisors, and other concerned parties and used by them to take various kinds of action in the context of the new information. Although at the University of Michigan we found the information helpful in relation to all students, it has proved especially useful for second language writers.

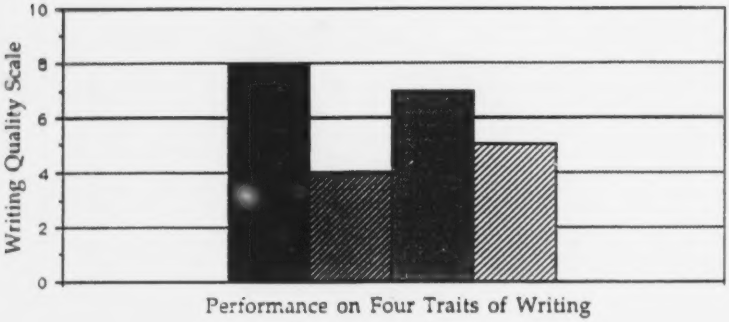
I have identified two types of profile which profile reporting can convey: the flat profile and the marked profile. In contrast to holistic scoring, where the reader who notices an unevenness of quality in the writing has no way to report this observation, and must somehow reconcile it as a single score, multiple trait scoring permits performance on different components or facets of writing to be assessed and reported. When the writing in any one sample looks rather similar from any perspective, with no visible peaks or troughs of skill, I call the set of scores on multiple traits which result a flat profile. When the writer shows no extreme variations in performance, as in the example in Figure 10 below, her writing performance may reasonably be expressed as a single score of "6" on a nine point scale without significant loss of information. This is what I mean by a "flat profile": the profile and the averaged score say basically the same thing. But sometimes, and more often with LEP writers for the reasons I discussed above, the writing quality looks rather different from some perspectives than from others. I call the set of scores which result from this unevenness a marked profile (Hamp-Lyons, 1987; Hamp-Lyons & Prochnow, 1989a). In the example in Figure 11, below, the resulting averaged score of "6" does not well describe what the reader sees in the writing, nor does it signal to the teacher what she should expect to encounter when working with this writer in class.

Figure 10
Flat Profile



827

Figure 11
Marked Profile



Knowing the information in the profile is particularly important in two types of cases. If a writer's overall performance puts her into the category of those who will receive special courses or other special services, by looking inside the information provided by the multiple trait instrument, that is by looking at the score profile, the writer, the class teacher, and the program administrator can make good decisions about which course offering or other kind of service would most help this individual writer make progress. Clearly, the provision of special services is particularly likely in cases of special needs students, LEP writers among them. Second, when a writer has generally sound writing skills but a particular weakness in just one area, a single number score would almost certainly fail to reflect the extremely marked aspect of writing performance but separate trait scores would reveal it. While the overall score may not indicate that the writer needs any special help, program administrators, college counselors, the teacher and the writer himself can see the unusual pattern and decide whether to take action about it. Here too second language users of English are likely to be in this category.

These applications to diagnosis and specialized services are the greatest benefits of multiple trait scoring. As the federal government continues to reduce the amount of funding for LEP and other students with special educational needs, yet hypes up the rhetoric about failing schools and this country's resulting decline in world markets at each opportunity, we need to find forms of assessment that will provide more information about LEP students' needs so that the limited resources available for services can be well spent. A multiple trait form of holistic writing assessment does this.

Figure 12 is an attempt to illustrate the ways that the information-rich data generated by a multiple trait type of holistic writing assessment, which uses profile reporting, may explain differences across classes. This type of detailed reporting across classes could answer some of the questions about unsatisfactory results from LEP-funded programs that have been caused by the inability of non-experts to understand the complexities of the problems LEP learners and their teachers face.

Figure 12
Multiple Trait Score Reporting (2) Program

EITHER numerical, e.g.:

	Class X				Class Y			
SCORE	T1	T2	T3	T4	T1	T2	T3	T4
6	5	6	2	2	3	4	7	8
5	6	5	6	4	5	4	7	9
4	5	8	9	8	8	6	6	6
3	8	5	9	8	7	9	5	4
2	5	6	3	5	5	6	3	2
1	1	0	1	5	2	1	2	1

OR by text description, e.g.:

Students in Class X were generally fairly competent in discovering and stating a problem, solution, and the connection between them, and their suggested problems tended to be reasonable and realistic. Students in the class tended to do less well in developing their ideas, with 13 of 30 scoring in the lower half of the range. It was noted that a number of the students in Class X have serious language problems, scoring low on the Language Control category: In particular, five students scored only 1 for Language Control, and five more scored only 2.
Students in Class Y (etc.)

In hypothetical Class X there are a number of LEP students, and their unfamiliarity with writing in English and with the full spectrum of the grammar of the language (I use the word in its linguistic rather than its lay sense here) shows up on the Language Control trait, where their performance contrasts strongly with that of the total group in Class Y, in which (also hypothetically) there are only three LEP students. Not only does the multiple trait report allow the identification of Language Control as the problem area, it also allows us to see that students in Class X as a whole are doing a good job on higher order cognitive skills such as problem solving, areas where they do not start from a disadvantage. If these data were combined and reported as though they came from a holistic scoring, all this information would be lost.

Salience and Wash back By "salience" I mean that the writing qualities evaluated, and the kinds of writing samples collected are those that have been found appropriate in the context where the assessment takes place. In the British Council writing test referred to above, for example, one writing task (known as the "convergent" task) called for students to read a text and prepare what was in effect a summary, selecting the correct factual content and putting it into a short text of their own, perhaps with graphical material, and using the appropriate vocabulary from the discipline. The multiple trait instrument I designed as a result of work with readers of this test contained the traits of content coverage, presentation format, linguistic features (especially register and lexis), and task fulfillment (see Appendix 3). This task is very unlike the writing task I have used as my example in this paper, where no special knowledge is assumed, no selection skills are called on, answers are expected to be all text, and a general vocabulary will suffice. Because the multiple trait procedure, like primary trait scoring, involves prompt specification and development as well as scoring and reader training, it is a prerequisite of a multiple trait instrument that there is a close match between the writing to be done and the skills and text facets to be evaluated. I argued earlier that all holistic writing assessment has positive wash back -- a positive effect on the teaching that goes on in the context leading up to the test. I believe that this positive wash back is greater for multiple trait forms of holistic writing assessment than any other. This comes from two primary sources: the careful, contextual test development which ensures congruence between teaching aims and testing values, and the provision of score consumers with descriptively informative and accurate test score information appropriate to their potential uses of it.

Improving on Multiple Trait Assessment

In developing writing assessment measures, I have always found myself in the situation of coming in after a good deal of water has flowed under the bridge, and trying to shore up the banks and reroute the waters through fertile lands. This means that certain desirable elements of excellence in a writing assessment are often not within practical reach. What are these? Some of them are commonly-accepted test characteristics that enhance accuracy of information by increasing the amount of information obtained. First, a basic principle of educational measurement is that the more items in a test the more reliable the information obtained will be: a writing test where the writers write several texts will provide more information about the range of the writer's skills in the contexts and traits that are salient. Second, all modern teachers of writing regret the limited amount of time available for writers to respond to prompts, since these speeded tests run counter to what we know about how successful writers write and to the philosophies of the "process" school of teaching writing. We would like more tasks, and more time: In the trade-off between time and task, there is some evidence (Livingstone, 1987; Hamp-Lyons & Henning, 1991) that LEP writers do not perform significantly differently when they have one hour to respond to a prompt than when they have only 30 minutes to respond to a prompt. And, when Michigan's State Writing Committee experimented with giving several days (a day and an hour for students in third, sixth and eighth grades) to respond to a writing prompt, there was no clear pattern of advantage for any of these below the eighth grade, where the longer led to higher scores. There is, however, considerable evidence (Reid, 1989; Hamp-Lyons & Prochnow,1990) that writers' performances vary considerably across task types. With a school-age population and an hour for a writing test, my preference would be, then, to shorten the time available for writing each task and have two tasks. A better option, of course, would be to increase the total amount of time and have two or more tasks with varying time limits. Another desirable element would be to have writing test data collected in small "bites" on several occasions rather than in the context of a stressful formal test situation. This is, of course, especially important with LEP students who may not be confident in their writing to begin with. Collecting a 30-minute sample once a week for three weeks gives the opportunity for different task types and different contexts, and also for the teachers to build the assessment into the curriculum, making it less intrusive and more educationally meaningful.

The two other elements on my "wish list" may not contribute to making writing assessment more accurate, although each is so poorly understood I don't think we can say that yet, but they would certainly contribute to making it more humanistic. First, it never fails to amaze me how little we know about what the test takers think about the tests, what they do when faced with a test, and I would like to see test design pay more attention to test takers' views and responses. As an example, we often hear it said that LEP students need longer to write on tests because their writing is not yet well-internalized. But we also often hear that LEP writers do less revising, and less global revising than advanced writers, and therefore are unlikely to take good advantage of additional test time, and that has been my own experience (Hamp-Lyons, 1990). But these two statements provide conflicting suggestions for test design. I don't think we can resolve these issues until we spend time in close observation of and conversation with LEP writers as they engage in the writing test event. And second, I think we should put some serious research effort into self assessment of writing. In my own classes, which typically contain both native and nonnative writers of English, I am becoming more and more courageous in introducing student self assessments into the assignment of end-of-course grades. I am finding that students who have taken a course with clear goals and pathways to achieving those goals finish the course with a very accurate internal sense of how good their writing is and where they need to improve, even though I never assign grades during the course. I find I rarely need to adjust the grade the student suggests for himself or herself by more than a half-grade: The exception seems to be in cases of long-term LEP residents who have made little progress in their English skills, typically because they have become absorbed into a local community of users of their first language and because they have avoided all situations where they might need to use English beyond the level they know they have already mastered. These students often greatly overestimate their writing competence. We have a great deal to learn about self-assessment, about what its benefits and problems are, but involving students in the assessment of their own competencies gives them a responsibility that may be repaid with greater understanding of their own strengths, weaknesses, and needs. It is when learners understand what they need, and take responsibility for filling their own needs, that they exercise the democratic citizenship rights we all believe in, that they move out from under the shadow of paternalism and condescension. We all, teachers and testers, must do all we can to help them make that move toward self determination.

Portfolio Assessment

A full consideration of portfolio assessment goes beyond the limits of this paper, but I must at least mention the rapid growth of interest in and practice of portfolio-based assessment of writing. I think the evidence is now strong that portfolio assessment will eventually become the preferred method for judging writing in many school and college contexts.

A portfolio is a collection of texts the writer has produced over a defined period of time to the specifications of a particular context. Portfolios, usually called "writing folders," have been used in formal assessment in England since the introduction of alternative school-leaving examinations in the early 1970s. Portfolios are used in many disciplines and at all school levels, but they seem to be especially appropriate both for the assessment of writing and for the assessment of the writing of LEP students. Individual high, junior high, and even elementary schools and school districts are using portfolios to monitor learning through the school year. Pittsburgh Public Schools have been developing portfolios in a range of subjects for some years, with a joint Rockefeller grant with ETS and Harvard Project Zero.

Having introduced an ambitious direct writing assessment in the late 1980s, California is now experimenting with portfolio assessment in consortia of schools. States such as Rhode Island are beginning to use portfolio assessment to obtain a picture of achievement in writing across the school system, and even a state with a very large school population such as Michigan has evaluated the need for and practicality of portfolio assessment at certain grade levels in order to obtain a "report card" of writing competencies statewide. Portfolio assessment is rapidly gaining ground at the college level too: at the University of Michigan, for example, they are used to assess exit competence from our pre-composition course (Condon & Hamp-Lyons, 1991; Hamp-Lyons & Condon, 1990), while schools such as Miami University of Ohio are beginning to use optional portfolios as part of entry assessment.

The portfolio usually does not contain writing produced under test conditions, although in some contexts such writing is also judged and considered in decisions such as whether exit competence standards have been reached. Some portfolios are simply a collection of responses to several essay test prompts, usually in different modes, while others incorporate drafts and other process data in addition to final products. The best portfolio assessments collect writing from different points over the course or year and take into account both growth and excellence. Such portfolios require students to include in their portfolio papers which have been revised over a period of time and to provide the original draft and all subsequent drafts. I know of no projects that explore portfolio assessment specifically as this applies to and affects nonnative writers at college level but, in the Michigan writing program exit assessment referred to above, we found that nonnative writers were more likely to be promoted to the next level than when promotion was based on impromptu writing alone. It seemed to us that the opportunities for multiple drafting, self-reflection, and receiving and responding to feedback implied by the portfolio mirror the reality of writing as it is taught these days and the ways students approach writing when it is required in their courses outside English class. Portfolios, because they contain several samples, and because they can be constructed so that texts written under different conditions are included, allow a more complex look at a complex activity, and are therefore generally considered to be more valid. Many problems, not only of reliability but also of the validity of readers' responses, training for portfolio reading, and others (Hamp-Lyons & Condon, 1990) remain to be solved, but the application of portfolio assessment in the ESL writing assessment context is an area that will repay attention in the next decade or less. I hope we will see many studies of portfolio assessment in LEP contexts before much longer.

Conclusion

My purpose in this paper has been to argue for direct, that is, holistic assessment of writing. Unlike some of my education colleagues, I **believe** in assessment, and I applaud President Bush's identification of assessment as a strategy for moving the country toward educational excellence. However, I agree with my colleagues Scott Enright and Mary Lou McCloskey, executive board members of TESOL, when they deplore the President's exclusion of teachers, the expert educators of the nation's youth, from primary input and participation in any of the national strategies including test design. I agree with them when they declare that "Our schools are already burdened by numerous standardized tests which put low-income and language minority students at a disadvantage" and that "we need new ways to recognize and utilize our students' genius, not new ways to label and sort students." (Enright & McCloskey, 1991, p.8). Most tests are based on a deficit model: they point out what the student cannot do, and special needs students are most in danger of suffering from the application of a deficit model to their educational needs. Multiple trait assessment in its most fully-developed form allows a description of both strengths and weaknesses, neither obliterating the other, an approach which holds great promise for LEP students.

Enright and McCloskey have noted that students with special needs are mentioned only once in AMERICA 2000, and in that reference they are referred to as "at risk". They note too that nowhere in the report is there any mention of the language minority population which makes up about 10 percent of the school-age population nationally. These are discouraging signs for those of us committed to the education of this group and to their integration as fully functioning citizens. Still more discouraging is the lack of reference to the underlying problems in this country, to poverty, malnourishment, lack of affordable child care and health care, to racism and alienation, to the abandonment of millions of women and children by their men and by the welfare system. Assessment is not a quick fix or a cheap fix: good assessment costs money. I think that holistic writing assessment, especially multiple trait assessment, offers a great value for money. But if our LEP children are sick, or homeless, or afraid; if our LEP adult students are unemployed, drug or alcohol addicted, or alienated by and from society, even the best assessments cannot help them.

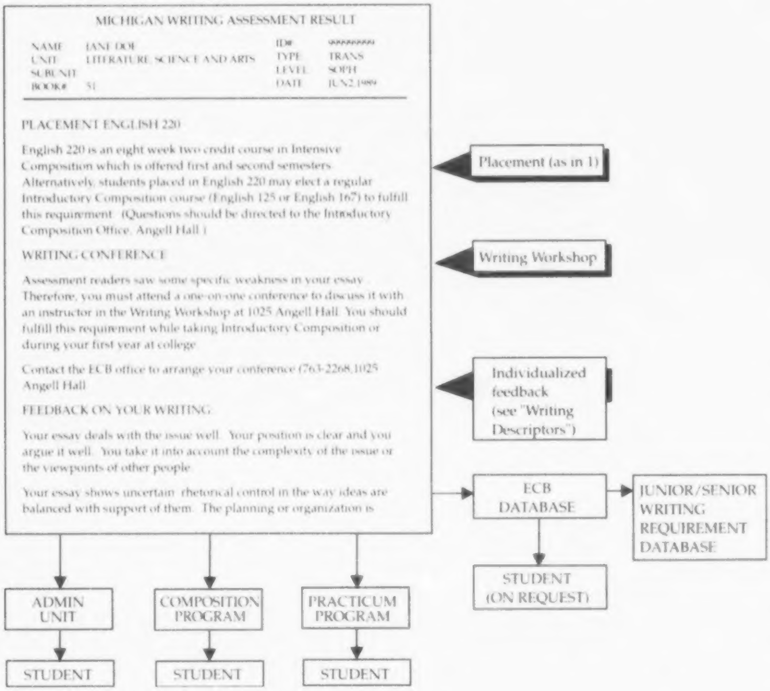
834

Appendix A
Michigan Writing Assessment Scoring Guide

English Composition Board: Criteria for Reading the Assessment

	Ideas and Arguments	Rhetorical Features	Language Control
6	The essay deals with the issues centrally and fully. The position is clear, and strongly and substantially argued. The complexity of the issues is treated seriously and the viewpoints of other people are taken into account very well.	The essay has rhetorical control at the highest level, showing unity and subtle management. Ideas are balanced with support and the whole essay shows strong control of organization appropriate to the content. Textual elements are well connected through logical or linguistic transitions and there is no repetition or redundancy.	The essay has excellent language control with elegance of diction and style. Grammatical structures and vocabulary are well-chosen to express the ideas and to carry out the intentions.
5	The essay deals with the issues well. The position is clear and substantial arguments are presented. The complexity of the issues or other viewpoints on them have been taken into account.	The essay shows strong rhetorical control and is well managed. Ideas are generally balanced with support and the whole essay shows good control of organization appropriate to the content. Textual elements are generally well connected although there may be occasional lack of rhetorical fluency: redundancy, repetition, or a missing transition.	The essay has strong language control and reads smoothly. Grammatical structures and vocabulary are generally well-chosen to express the ideas and to carry out the intentions.
4	The essay talks about the issues but could be better focussed or developed. The position is thoughtful but could be clearer or the arguments could have more substance. Repetition or inconsistency may occur occasionally. The writer has clearly tried to take the complexity of the issues or viewpoints on them into account.	The essay shows acceptable rhetorical control and is generally managed fairly well. Much of the time ideas are balanced with support, and the organization is appropriate to the content. There is evidence of planning and the parts of the essay are usually adequately connected, although there are some instances of lack of rhetorical fluency.	The essay has good language control although it lacks fluidity. The grammatical structures used and the vocabulary chosen are able to express the ideas and carry the meaning quite well, although readers notice occasional language errors.
3	The essay considers the issues but tends to rely on opinions or claims without the substance of evidence. The essay may be repetitive or inconsistent; the position needs to be clearer or the arguments need to be more convincing. If there is an attempt to account for the complexity of the issues or other viewpoints this is not fully controlled and only partly successful.	The essay has uncertain rhetorical control and is generally not very well managed. The organization may be adequate to the content, but ideas are not always balanced with support. Failures of rhetorical fluency are noticeable although there seems to have been an attempt at planning and some transitions are successful.	The essay has language control which is acceptable but limited. Although the grammatical structures used and the vocabulary chosen express the ideas and carry the meaning adequately, readers are aware of language errors or limited choice of language forms.
2	The essay talks generally about the topic but does not come to grips with ideas about it, raising superficial arguments or moving from one point to another without developing any fully. Other viewpoints are not given any serious attention.	The essay lacks rhetorical control most of the time, and the overall shape of the essay is hard to recognize. Ideas are generally not balanced with evidence, and the lack of an organizing principle is a problem. Transitions across and within sentences are attempted with only occasional success.	The essay has rather weak language control. Although the grammatical structures used and vocabulary chosen express the ideas and carry the meaning most of the time, readers are troubled by language errors or limited choice of language forms.
1	The essay does not develop or support an argument about the topic, although it may "talk about" the topic.	The essay demonstrates little rhetorical control. There is little evidence of planning or organization, and the parts of the essay are poorly connected.	The essay demonstrates little language control. Language errors and restricted choice of language forms are so noticeable that readers are seriously distracted by them.

Appendix B
Michigan Writing Assessment Scoring Report



MICHIGAN WRITING ASSESSMENT RESULT

NAME JANE DOE
UNIT LITERATURE, SCIENCE AND ARTS
SUBUNIT
BOOK# 51
ID# 999999999
TYPE TRANS
LEVEL SOPH
DATE JUN2,1989

PLACEMENT ENGLISH 220

English 220 is an eight week two credit course in Intensive Composition which is offered first and second semesters. Alternatively, students placed in English 220 may elect a regular Introductory Composition course (English 125 or English 167) to fulfill this requirement. (Questions should be directed to the Introductory Composition Office, Angell Hall.)

WRITING CONFERENCE

Assessment readers saw some specific weakness in your essay. Therefore, you must attend a one-on-one conference to discuss it with an instructor in the Writing Workshop at 1025 Angell Hall. You should fulfill this requirement while taking Introductory Composition or during your first year at college.

Contact the ECB office to arrange your conference (763-2268,1025 Angell Hall

FEEDBACK ON YOUR WRITING

Your essay deals with the issue well. Your position is clear and you argue it well. You take it into account the complexity of the issue or the viewpoints of other people.

Your essay shows uncertain rhetorical control in the way ideas are balanced with support of them. The planning or organization is

BEST COPY AVAILABLE

Appendix C
British Council ELTS M2 Writing Sub-test: Convergent Task Scoring

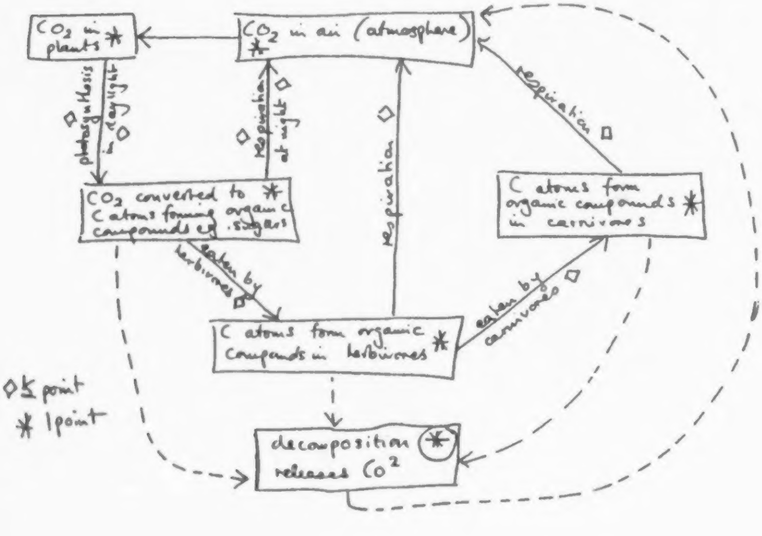
MARKING SUB-SCALES FOR QUESTION 1

CONTENT COVERAGE

THERE IS NO SUB SCALE FOR CONTENT COVERAGE
MARKERS SHOULD REFER DIRECTLY TO THE PROTOCOLS

SAMPLE:
LIFE SCIENCES PROTOCOL 1

(Questions 1 of Versions 4, 5, 6)



CO2 in plants *
CO2 in air (atmosphere) *
photosynthesis ◇
in daylight ◇
respiration ◇
at night ◇
respiration ◇
respiration ◇
CO2 converted to C atoms forming organic compounds eg. sugars *
C atoms form organic compounds in carnivores *
eaten by herbivores ◇
C atoms form organic compounds in herbivores *
eaten by carnivores ◇
decomposition releases CO2 *
◇ ½ point
* 1point

Points

- - - → • this point is not overtly stated in the text. Candidates may receive 1 point if they include it and have not got all the other Content points.

Presentation Format
Flow diagram.
Outline: all stages must be clearly sequenced.

Appendix C (Continued)

SUB-SCALE for PRESENTATION FORMAT

BAND	DESCRIPTOR
9	The most suitable presentation format is used. It is applied in a way that shows full mastery of it in presenting main points and details.
7	A suitable presentation format is used. The format is applied effectively in general, although one or two inaccuracies in the application of the format to the details may be observed.
5	EITHER A suitable presentation format is used, but it is not applied effectively in the presentation of the information. OR An unsuitable presentation format is used, but it is applied effectively in the presentation of the information.
3	An unsuitable presentation format is used. There are many inaccuracies in the application of the format to the main points and details.
1	No evidence of control over a comprehensible presentation format can be observed.

Appendix C (Continued)

SUB-SCALE for TASK FULFILLMENT

BAND	DESCRIPTOR
9	The overall impression is of a set of notes which fulfills the task fully, clearly and with complete subject command and language control. No irrelevant or inaccurate information is included.
8	The overall impression is of a set of notes which fulfills the task fully, clearly, and with good subject command and linguistic control. No, or very little, irrelevant or inaccurate information is included.
7	The overall impression is of a satisfactory answer which fulfills the task with only occasional, minor, flaws in the subject or language control. Some irrelevant or inaccurate information may have been included, but the clarity of the answer makes it possible to ignore this.
6	The overall impression is of a mainly satisfactory answer although there are some minor flaws of subject or language which detract from the fulfillment of the task. Some irrelevant or inaccurate information may have been included, but this does not seriously impinge on the presentation of the essential material.
5	The overall impression is of an adequate answer, but failure to include some essential information, uncertainly in presenting the notes, language hesitancies, or the inclusion of irrelevant or inaccurate information detract from the satisfactory fulfillment of the task.
4	The overall impression is of an answer which, although it makes a valid attempt to fulfill the task, is too flawed by problems such as lack of information, an inappropriate or unclear approach to note-making, inappropriate transfer from the input text or task, irrelevance, inaccuracy or language weakness to be considered adequate.
3	The overall impression is of an answer which attempts the task but is so seriously flawed in several areas (as listed in band 4) that it does not approach a fulfillment of the task.
2	The seriousness of the flaws in this answer make it impossible to judge it in relation to the task set.
1	A true non-writer who has produced no assessable notes, either because of evident lack of command or because the answer has been lifted wholly or almost wholly from the input text or task (please note which category on the front of the candidate's answer paper).

Appendix C (Continued)

SUB-SCALE for LINGUISTIC FEATURES

BAND	DESCRIPTOR
9	There are no errors or omission in the candidate's application of conventions of register. Key lexis, if appropriate, is present and used correctly. No errors of accuracy or appropriacy in the candidate's linguistic control.
8	There are no errors in the candidate's application of conventions of register but the marker may be aware of certain features of register which would have been appropriate but which are not present. Key lexis, if appropriate , is present and used correctly. There is no inappropriate transfer of key lexis from the input text or task. There are no significant errors of accuracy or appropriacy in the candidate's linguistic control.
7	There may be one or two errors in the candidate's application of conventions of register, and/or the marker may be aware of certain features of register which would have been appropriate but which are not present. The candidate may fail to transfer key lexis when appropriate, but there is no inappropriate transfer of key lexis from the input text. There are occasional minor errors of accuracy or/and appropriacy in the candidate's linguistic control.
6	Several errors are noted in the candidate's application of conventions of register. The marker may be aware of restricted range of register features, or of a failure to transfer appropriate key lexis from the input text, but key lexis is not transferred inappropriately. There are a number of errors or linguistic accuracy and a limited ability to manipulate the linguistic system appropriately.
5	Several errors are noted in the candidate's application of register of conventions. The marker is aware of a restricted range of register features and of a failure to transfer key lexis when appropriate. One or two key lexical items may be transferred inappropriately. Linguistic errors of accuracy or appropriacy intrude frequently.
4	The marker notes a lack of overall command of appropriate register, although one or two appropriate features may be present. The candidate does not transfer key lexis when appropriate. One or two key lexical items may be transferred inappropriately. The control of the linguistic system is generally inadequate. The effect of these failures and omissions is to make retrieval of the information difficult.

Note

[1] If 10 percent of scores received a third score, for example, in the formula K would hypothetically be 2.10 and attenuated reliability would be enhanced: however, a third reader would only be needed in 10 percent of cases if the first two readings were quite unreliable or the standard for a discrepant score very stringent. Standards for recognizing a score as discrepant vary considerably: the TOEFL Program's TWE requires third readings on the basis of a two-point discrepancy on a six scale (33 percent discrepancy criterion), the MELAB uses a two-point discrepancy criterion on a nine-point scale (22 percent discrepancy criterion), and the Michigan Writing Assessment uses a six-point discrepancy on a thirty-six point scale (16.5 percent discrepancy criterion).

References

U.S. Department of Education (1991). AMERICA 2000: An Education Strategy. Washington, DC, Author.

Anastasi, A. (1982). Psychological Testing (5th ed.). London: Collier Macmillan.

Bachman, L. (1990). Fundamental Considerations in Language Testing. Cambridge: Cambridge University Press.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing. Research in the Teaching of English, 18, 65-81.

Condon, W. C., & Hamp-Lyons, L. (1991). Introducing a portfolio-based writing assessment: progress through problems. In P. Belanoff, & M. Dickson (eds.), Portfolio grading: Process and product. Portsmouth, NH: Boynton/Cook.

Cooper, C. R. (1977). Holistic evaluation of writing. In C. R. Cooper, & L. Odell (eds.), Evaluating writing: Describing, measuring, judging. (pp. 3-32). Urbana, Il: NCTE.

Cronbach, L. (1949). Essentials of psychological testing. New York: Harper & Brothers.

De Jong, J. H. A., & Henning, G. (1990, March). Test dimensionality in relation to student proficiency. Paper presented at the Twelfth Annual Language Testing Research Colloquium, San Francisco.

Enright, S. & McCloskey, M.L. (1991, Aug/Sep). America 2000: Two TESOL members respond. TESOL Matters, 1, (4), 1-8.

Godshalk, F. I., Swineford, F., & Coffman, W. E. (1966). The measurement of writing ability, ETS Research Monograph, 6. Princeton, NJ: Educational Testing Service.

Hamp-Lyons, L. (1984; revised 1986). Assessment Guide for M2 Writing. London: The British Council.

Hamp-Lyons, L. (1987). Performance profiles for academic writing. In K. Bailey, R. Clifford, & T. Dale (eds.), Language Testing Research: Papers from the Ninth Annual Language Testing Research Colloquium. Monterey, CA. : Defense Language Institute.

Hamp-Lyons, L. (1990). Essay test strategies and cultural diversity. Pragmatic failure, pragmatic accommodation, and the definition of excellence. [ERIC ED 07-90-019.]

Hamp-Lyons, L., & Condon, W. C. (1990, Nov). Readers' responses to portfolios. Paper presented at the Eight Annual National Testing Network in Writing, New York.

Hamp-Lyons, L., & Henning, G. (1991). Communicative writing profiles: An investigation of the transferability of a multiple-trait scoring instrument across ESL writing assessment contexts. Language Learning, 41, (3), 337-373.

Hamp-Lyons, L., & Prochnow, S. (1989, March). Person dimensionality, person ability and item difficulty in writing. Paper presented at the Eleventh Annual Language Testing Research Colloquium, San Antonio, March.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). Testing ESL composition: A practical approach. Rowley, MA: Newbury House.

Livingstone, S. (1987, April). The effects of time limits on the quality of student-written essays. Paper presented at the annual meeting of the American Educational Research Association, New York.

McKenna, E. & Carlisle, R. (1991). Placement of ESL/EFL undergraduate writers in college-level mainstream writing programs. In L. Hamp-Lyons (Ed.), Assessing Second Language Writing in Academic Contexts. Norwood, N.J.: Ablex.

Purves, A. C. (1984). In search of an internationally-valid scheme for scoring compositions. College Composition and Communication, 35(4), 426-439.

Reid, J. (1989). Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (Ed.), Second Language Writing: Research insights for the classroom. (pp.191-211). New York: Cambridge University Press.

Weir, C. (1983). Identifying the language problems of overseas students in tertiary education in the United Kingdom. Unpub. doctoral dissertation, University of London.

White, E. (1985). Teaching and Assessing Writing. San Francisco: Jossey-Bass.

Wiseman, S. (1949). The marking of English compositions in grammar school selection. British Journal of Educational Psychology, 19(3), 200-209.

Response to Liz Hamp-Lyons' Presentation

Denise McKeon
National Clearinghouse for Bilingual
Education, Washington, DC

It's a pleasure for me to be here today, and it's also a pleasure to respond to Liz Hamp-Lyons' paper. As a former bilingual and ESL classroom teacher, I have never been a very big fan of assessment and, in particular, standardized assessment. It's not that I believe that we don't need to know how students are doing. It's not that I believe that assessment is inherently bad. It's just that the assessments that have traditionally been used with limited English proficient students, such as standardized multiple-choice tests, had a way of neglecting to show all the things that my kids could do and all the things that they had learned.

In addition, testing always seemed to take away valuable time and resources from instruction and never seemed to give much back. What I always wanted from assessment was some kind of measure that would point me in the right direction, instructionally, with my students; something that would provide me and the students with some guidance as to how to move closer toward that illusive goal of becoming proficient in English. Liz Hamp-Lyons has shown me that there may be hope -- that assessment has really come a long way. She documents the move toward holistic assessment quite eloquently and echoes the concerns that most teachers and responsible test developers have expressed about the testing processes used in assessing writing skills.

I should point out that this movement away from multiple-choice tests of discrete writing skills is linked to a national movement, which parallels the call for development of national standards and school reform, as you've heard enumerable times today, and I'm sure, will continue to hear throughout the course of this program. While many from the school reform movement are calling for a national test, some of those charged with the responsibility of designing and implementing assessment, such as the New Standards project, are, thankfully, exploring ways of making testing more representative of what students really need to know and learn. They are looking for ways to put the instructional cart back behind the testing horse, having curriculum drive instruction, rather than the other way around. Holistic writing assessment is one component of this responsible testing movement.

As Liz pointed out in her paper, there have been some concerns expressed about the reliability of scoring such holistic assessments.

844

While the amount of information that we have with regard to scoring reliability in relation to LEP students, is quite small, recent evidence of scoring reliability with mainstream populations suggests that this might not be the problem that it has previously been thought to be.

The New Standards project, for example, convened a meeting of more than one hundred elementary, middle school, and high school teachers and educators from five states in July of this year. These teachers and educators, who were conducting direct writing assessments in their own states, met to score each other's sample student papers using his or her own state's rubrics for scoring. The purpose of this activity was to examine whether it was possible to calibrate or compare the results from prompts developed by different states and scoring rubrics developed by different states. The results were astounding. Cross state inter-scorer reliabilities in the range of .81 to .87 were obtained, leading Dan Resnick, one of those involved with the project to remark, "it appears that there are conditions under which human judgment can be trusted."

What seems to be needed in holistic writing assessment of ESL students is both the development of standards and scoring rubrics that are sensitive to students acquiring English as a second language, as well as some accompanying professional development of teachers and educators, which will nurture the type of trained professional judgment that has been shown to be so powerful. There also seems to be some need for ESL teachers and other educators to discuss what it means for an LEP student to be a 3 as opposed to a 4 on a writing test. There appears to be a need for teachers and other educators to discuss what it means to be a proficient writer of English as a second language. Is it the same as being a proficient writer of English? Would holistic writing assessment of LEP students in K to 12, for example, be tied to some measure of exit from ESL instructional programs? Teachers who have had little experience with LEP learners might find the unevenness of their writing surprising. Scoring rubrics that help to alert teachers to the types of unevenness that are to be expected across certain levels of language development could help to guide those teachers in making more accurate assessments of students' abilities.

Let me just say a word here. I have been talking with some people that have been working in writing projects in the Northern Virginia area. They've noticed some very interesting things about the way certain students respond to these types of test taking circumstances. When certain groups of LEP students, for example, who really do quite well in class on a regular basis, get into these test taking situations, they are very afraid of being wrong. They tend to hold back and, in holding back, they tend not to perform as well as they could have if they would have gone ahead and taken the risk, because risk taking happens to be one particular point of a given scoring rubric. It's a catch-22 which is very interesting.

The development of scoring rubrics becomes even more critical when students who are in the beginning stages of writing are encouraged to use their native languages in school programs. There has been very little work done in determining what certain levels of writing look like in languages other than English at the K to 12 level.

Teacher input into the development of such scoring rubrics is a source of professional development in and of itself. The more experience teachers and other potential scorers have, not only with varieties of ESL or L-1 writing, but also with how those varieties fit against some scale of second language writing, the more they will be able to rate those types of writing discerningly.

There is an additional issues that needs to be raised here, that of the amount of experience that teachers, themselves, have with writing. One question remains to be answered: What is the relationship of scoring patterns of teachers to their own writing experience and competence? In other words, do teachers, who write on a regular basis, score students differently than teachers who do not? Do teachers, who write well, score students differently than teachers who do not? To date, those who have been involved with assessment of holistic writing are those who have had great interest in writing. They believe writing is a valuable skill. They believe in practicing that skill though process approaches and conferencing, and they believe in the use of holistic measures as a viable assessment system. This is a very important feature of what has occurred in holistic language assessment to date.

I am not arguing with any of this. What I am suggesting, however, is that, as more and more teachers and educators become involved in such testing, many of those who become involved will be as crazy about whole language, process writing, and holistic assessment. What will happen as those less enchanted teachers are asked to administer and score holistic language assessments? It would seem important to compare scoring results between those who are "experts with writing" and those who are, for want of a better term, "novices." It would also seem equally important to compare scoring results of those who are fans of holistic language approaches and holistic assessment, and those who are not.

Portfolio assessment, which Liz talked about a good deal in her paper, is one area of writing assessment which has received a great deal of attention and shows great promise for judging writing in many schools and college contexts. The portfolio provides an opportunity for teachers to view multiple samples of student work including work that has undergone revision. One important benefit of portfolio assessment is that both teachers and students begin to see the evaluation process as one which involves growth, rather than as one which is an immutable static measure of competence at some point in time. Portfolio assessment allows teachers and students to engage in collaborative examination, examination that provides students with some measure of control in the examination process. What is necessary to determine is how certain pieces, which contribute to the portfolio, are selected. Are the pieces selected by the teacher alone? By the student? What types of writing are determined to be necessary for inclusion? If portfolio assessment is to be used as a representative measure of student work, care must be taken to be as inclusive as possible of all the types of writing that a student is being asked to learn and practice as part of instruction.

Given the paradigm shift that has occurred in K-12 ESL instruction in recent years, this means attending to the emergence and presence of content-based ESL. As more and more programs begin to introduce content-based ESL or sheltered English, the presence of such subject matter must also begin to be addressed in portfolio assessment. Just as writing across the curriculum becomes an important part of content-based ESL classes, it must also be examined through portfolio assessment. The examination of student writing by both trained ESL and content teachers could help to build instructional bridges that result in more meaningful instruction for LEP students.

Another benefit of portfolio assessment deals with the notion of eliciting student work in naturalistic settings. These naturalistic settings allow three things to occur. Student work can be produced under "normal" classroom circumstances, in other words, on a non-timed basis. Student work can be seen as evolving, and data can be collected which reflects students' thinking about the nature of writing. The inclusion of multiple drafts of a particular piece of work allows both teacher and student to reflect on the effect of the instructional and learning process over time.

Since one of the ultimate goals of writing is to produce writers who can self-edit and self-evaluate, the representation of this process in the portfolio is critical. The naturalistic setting in which work for inclusion in portfolios is developed is further enhanced by the underlying assumption that conferencing is an important part of holistic writing approaches. Through conferencing, portfolios and the work which they contain become a reason for talking and thinking about the ways in which language and content interact.

One of the most important benefits derived from portfolio assessment by way of conferencing is the ability to explore metacognitive aspects of student writing. Students can and should be asked ques-

847

tions such as "how do you know a piece is getting better?, how can you tell that someone is a good writer?, what kinds of things do you usually do to make a story more interesting?" These expressions of student intent and understanding provide important clues about what students know and understand about writing. Additionally, they offer the teacher insight into students' conceptions of what writing is, further providing opportunities for teachable moments.

One additional use of portfolios may be to use them to train student judges of writing. One of the biggest drawbacks in writing has traditionally been that students rarely get to read the work of other students or, for that matter, the teacher. Portfolios provide an opportunity for students to interact with the work of others and to serve as editors to others, by offering suggestions that may ultimately serve as self-instruction. Perhaps the biggest benefit to be derived from portfolio assessment and other types of holistic writing assessment is that they may actually affect a change in how classes designed for limited English proficient students are taught.

While many ESL and bilingual classes have moved to whole language approaches, there are still many places where whole language is not readily accepted. This raises the question of whether holistic language assessment is a viable approach to use with those more traditionally taught ESL classes. While I generally deplore the notion that tests may drive instruction, a move toward holistic language assessment may actually have the effect of changing the way in which instruction gets delivered. You can't perform well on a writing test if you haven't had any experience with writing in class. This fact alone may induce certain districts and teachers who are reluctant about holistic writing approaches to try them.

Thus, performance based assessments may eventually nudge schools away from the reductionist "kill and drill" form of instruction to instruction which enables students to perform well, not only for the tests, but in real life. If nothing else, holistic language assessment will have assisted the processes of teaching and learning greatly if only this is accomplished.

Response to Liz Hamp-Lyons' Presentation

Joy Kreeft Peyton
Center for Applied Linguistics, Washington, DC

There is a great deal to celebrate about this paper, so my response begins with celebration of many of its points. I follow that with some comments about areas in which I think we need to push further, and I close with some questions that still remain for me and probably for most of us.

Celebration

It was extremely heartening, reading this paper and listening to Liz talk, to realize the progress we have made in our thinking about what writing is and how we can best assess its quality. A paper about writing assessment written 10 years ago might have begun with extensive discussion of what Liz calls objective tests (which could also be called indirect tests, since they don't assess writing itself but related sub-skills) and then as a wish, suggestion, or afterthought move to a brief discussion of assessing actual writing samples. This paper begins with the recognition that holistic scoring of actual pieces of writing is the only way writing can be assessed, offers a well-developed and much-needed critique of this approach, and moves us along further with a description of multiple trait scoring. For me, this reflects a great and long-in-coming leap forward in our thinking about writing and its assessment, even though, as Liz acknowledges, direct writing assessment is still a young field, and there is still a lot more work to do.

It is also heartening to realize how far we have come in understanding the importance of content, task, and context in the quality of writing products, and the need to take those into consideration when designing an assessment. We now know that a valid writing assessment must begin long before testing actually takes place, with a needs assessment to determine what the writing context and teaching aims are and what qualities of writing are desired. For far too long, we have designed, scored, and accepted the results of decontextualized writing tests, and we have had very little idea of what actually went on in the programs and classes involved or even what the participants were actually trying to accomplish.

In her "wish list" at the end of the paper, Liz mentions a number of ways that writing assessment might be improved even more:

- Involving teachers in test development and scoring.

- Providing for multiple, revised drafts.

- Collecting writing regularly during the year from different contexts and types of tasks, as part of instruction and not separate from it.

- Including portfolio assessment as an option for even large-scale assessments. (Liz says this will eventually become the preferred method for assessing writing, and I hope it does.)

- Observing students as they compose, to better understand how they approach the tests we design. (The methods for this are already well-developed, through the writing protocol research, and computer programs allow us to do it unobtrusively without imposing on students and without asking them to talk while they compose. For example, Recording WordStar, developed at the University of Minnesota (cf. Bridwell, Sirc, & Brooke, 1985), plays back a student's composing session, and the student and researcher can talk about what the student did and why.

Carmen mentioned this morning that this conference would be helping to set a research agenda, and I think these items on the wish list should be part of that agenda.

That these approaches are already being tried on a small scale in a number of places is another indication of the progress we are making, and I hope that as we continue to think about writing assessment, they will move to the beginning of our papers and the forefront of our thinking and research.

Finally, I celebrate something that Liz laments--the genuine and truly educational activities mentioned early in the paper: taking a field trip to the pond, carrying out an experiment on specific gravity, writing a poem about an important experience, and so on. Liz mentions, for example, that in the school district where her first grade son attends, he didn't take even one field trip during the year, and that if this is a trend in education, it's a lamentable one, and I agree. Although Liz bemoans the absence of these kinds of activities in our schools, they are precisely the kinds of activities now advocated by leading teachers and researchers across the country. They may not yet be hailed in discussions of educational goals at the national level and they may not have reached all school districts (they evidently haven't reached the district in which Liz's first grade son goes to school), but they are slowly gaining recognition and respect, and I believe they will eventually prevail over skill and drill exercises to help students pass some standardized test.

Comments

There are a couple of areas where I would like to see us push further:

First, in the discussion of whether students need more time to complete a writing task, presumably to allow them to draft and revise; Liz mentions research finding that limited English proficient writers do very little revising and don't make good use of additional test time anyway. Therefore, they don't perform differently when given 30 minutes, an hour, or even several days to write. I believe the reason for this is that students have not been taught how to revise. They are so accustomed to submitting first drafts as final products to be evaluated that they don't know what to do with time for revision when they have it. If we want students to benefit from time for producing multiple, revised drafts, we need to teach them how to draft and revise. Until that process becomes a regular part of instruction, we can't expect to see it in assessments.

Second, we may be asking too much of large-scale writing assessments, designed primarily to determine how schools across the nation are doing, to evaluate individual programs, or to make decisions about student acceptance or placement when we ask that they not only yield numbers that can be compared but that they also give correction and feedback to writers. I wholeheartedly agree that writers need "sensitive and detailed feedback on their writing," but no amount of score detail can provide that. Multiple scores on well-defined traits can certainly give a rough indication of where a student is strong or weak and needs to work more, but they cannot replace thoughtful qualitative response to writing. Decisions about how many and what traits to score, whether or not to weight the scores, and whether or not to report the full score profile or only the composite score are all important at the administrative or policy level, but they provide little help to a student working on his or her writing. In the quest for the most descriptive test scores, we need to assure that those scores don't replace actual responses that are relevant and meaningful to individual learners. Someone still needs to react to students' text with text.

Questions

Finally, I have some questions that I don't think any of us have answers to at this point.

First, I don't know how national or even district-wide writing assessments can be very context-specific. Student characteristics, teacher goals, and program exit criteria can be as diverse and numerous as the teachers, programs, and classrooms themselves, and I don't see how a district or nationally developed assessment can possibly be sensitive to that diversity. The description and scoring of particular traits of writing seem extremely useful within a program or classroom, but can we expect agreement on which traits are important on any broader scale than that?

Second, I think we need to continually question what is the match between what we do and assess in school and the challenges that actually face students when they leave our programs. Whether we use "objective" tests, holistic scoring, multiple trait scoring, or writing portfolios, we still run the risk of focusing solely on school-based writing, which may have little relation to the literacy tasks demanded in the work place (see Harste & Mikulecky, 1984; Mikulecky, 1990). In deciding what students need to be able to do and, therefore what we will assess, we need to be sensitive and responsive to the continually changing situations those students will enter when they leave our programs.

For example, our discussions of writing assessment, whatever the format, revolve almost exclusively around the production of extended, usually expository text, by one author working alone. With the increasing emphasis on collaborative work both in school and in the workplace, is solitary text production really what students will do or need to be able to do? Or is this simply a vestige of our academic tradition, which no longer reflects the way we or our students actually work -- in collaboration with others? In future papers on writing assessment, I would like to see serious attention paid to the implications of collaborative writing practices.

Third, what do the students themselves want and feel they need to learn? Hunter and Harman (1979; cited in Wiley, 1991) note that assessment measures are not negotiated with those tested, but imposed largely by middle-class educators. Involving teachers in the assessment process or studying what students do with the tasks we design may be only first steps. In some portfolio assessments students not only select which writing pieces to include but also critique their own writing and prepare the portfolio for assessment. Is it possible to involve them even more, even possibly in deciding the kinds of writing they will do and helping to establish the evaluation criteria? Especially in programs for adults, it seems that our writing contexts and tasks need to encompass the contexts and tasks in which the students also find value.

Conclusion

We have come a long way in our thinking about writing and its assessment; but there is still more to do, and there always will be more to do, if we are going to be truly responsive to students' learning needs and desires and to society's changing needs for a literate population. Maybe I'm overly optimistic, but I believe that, as we continue to grapple together with that challenge of the linguistic and cultural diversity now prevalent in our schools and as we test and research new approaches to teaching and assessment now available to us, we will return to an understanding of "education" not as mastery of a set of specific skills, but rather, as Liz suggests, as preparation for life and citizenship and for social and moral responsibility.

References

Bridwell, L., Sirc, G., & Brooke, R. (1985). Revising and computing: Case studies of student writers. In S. W. Freedman (Ed.), The acquisition of written language: Response and revision (pp. 172-194). Norwood, NJ: Ablex.

Harste, J.C., & Mikulecky, L.J. (1984). The context of literacy in our society. In A.C. Purves & O. Niles (Eds.), Becoming readers in a complex society (pp. 47-78). Chicago, IL: The University of Chicago Press.

Hunter, C., & Harman, D. (1979). Adult illiteracy in the United States. New York: McGraw-Hill.

Mikulecky, L.J. (1990). Literacy for what purpose? In R.L. Venezky, D.A. Wagner, & B.S. Ciliberi (Eds.), Toward defining literacy (pp. 24-34). Newark, DE: International Reading Association.

Wiley T. (1991). Measuring the nation's literacy: Important considerations. Washington, DC: National Clearinghouse on Literacy Education.

A Superintendent's Evaluation of Teacher Education Reform

Transforming American Education: Making It Work

Peter J. Negroni
Springfield Public Schools, Massachusetts

American education like the U.S.S.R. is undergoing the most dramatic self-analysis since we decided early in our history that education would be available to everyone in this country. However, it seems that communism is easier to change than the present framework of American education.

In America, we became used to identifying the problems and then simply using good old American ingenuity, resoluteness, stick to it, determination, and get the job done. Well, we are finding that the old ways simply do not work anymore. Thus, America is on a mission, a search to improve the way it does everything so it can stay globally competitive with the rest of the world; and so it is with education.

This paper will take its reader through the course taken by this country to improve its public education system in response to the issue of global competitiveness. It will describe the nation's attempts at reform on public schooling since 1983 and chart teacher education reform within the broad reform context.

The paper will then share with you what has happened in the Springfield Public Schools and how the reforms can take hold in transforming schools.

For American public education, it would seem 1983 was the year that we discovered something was wrong with our schools. The report "A Nation at Risk" provides a broad set of recommendations for reforming our public schools. However, it is important to note that America has had other comprehensive reports that have called for sweeping reform in public education prior to 1983. Preceding "A Nation at Risk" were The Report of the Committee of Ten in 1983, the American High School Today in 1959, and the Cardinal Principals of Secondary Education in 1918. All of these, while different in intent and content, were dramatic efforts, to reform public education. Thus, to understand the present reform efforts, they must be viewed within the broader context of American educational reform.

854

The recommendations from "A Nation at Risk" are summarized in this table.

Recommendations from A Nation at Risk

I. Content

A. High School graduation requirements raised five new basics:

1. Four years of English: extended reading and writing skills and knowledge of our literacy heritage.

2. Three years of math:

    a. Higher-order mathematics such as geometry, algebra, and statistics.
    b. Estimation, approximation, measurement, and accuracy testing.
    c. A curriculum for those not planning college immediately.

3. Three years of science:

    a. Higher-order sciences, scientific reasoning, and inquiry.
    b. Application of scientific knowledge and technology.

4. Three years of social studies:

    a. Studies of selves and others in the continuum of time and culture.
    b. Understand social, economic, and political systems.

5. A half-year of computer science:

    a. Basic computer literacy and use of computers in other subjects.
    b. Comprehension of electronics and related technologies.

6. For the college-bound, 2 years of foreign language in high school is strongly recommended, in addition to 4-6 years of such study in the elementary grades (p.p. 24 and 26).

B. Upgrade elementary curriculum -- foreign language, English development in writing, problem-solving skills, science, social studies, and the arts.

C. Outside experts to improve and disseminate quality curricular materials: Evidence of text quality and currency from publishers.

II. Standards and Expectations

A. All educational institutions to adopt more rigorous academic standards: Grades to be indicators of achievement.

B. Standardized tests of achievement at transition points.

III. Time

A. More learning time: efficient time use, longer day, or longer year.

   1. More homework and instruction for study skills

   2. Districts to consider seven-hour days and 200- to 220- day schools years.

   3. Efficient management of the school day and class organization.

   4. The strengthening of attendance incentives and sanctions.

   5. Reduction of administrative and discipline burdens, and intrusion on teachers.

IV. Teaching

A. Improve preparation for and desirability of teaching

   1. Higher standards for incoming teachers; judge programs by quality of graduates

   2. Competitive, market-sensitive, and performance- based salaries; career decisions based on evaluation.

   3. Career ladders and 11-month contract

   4. Alternative credentialing, grants, and loans to attract teachers

   5. Master teachers' plan programs for probationary teaching and supervision

V. Citizen and Federal Involvement and Fiscal Support

A. Citizens oversee reform and provide financial support

B. Administrative and legislative officials provide stability and finance for reforms.

C. Federal government identifies national interest, provides leadership, and supports states and local districts.

> Source: National Commission on Excellence in Education, A Nation at Risk: The imperative for educational reform (Washington, DC: U.S. Government Printing Office.)

As one can see from this table, the recommendations are divided into five major areas:

1. Content
2. Standards and Expectations
3. Time
4. Teaching
5. Citizen and Federal Involvement and Fiscal Support.

It is important to note that this report, while it has been the cornerstone of the reform effort, has undergone a myriad of additions as a result of the proliferation of additional reports and studies undertaken since "A Nation at Risk." This very well may be the deciding difference between this reform period in American public education and the others mentioned above.

The initial reaction of states to "A Nation at Risk" centered around considering over 1000 pieces of legislation concerning teachers and teaching from 1983 to 1988 (Darling - Hammond and Berry 1988). The concentration during this period of time was to emphasize state driven and state mandated changes to reforms. These included increased time in school, more courses to graduate from high school, defined curricula, and promotional standards. These reforms were very evident during the author's superintendency in New York City between 1978 and 1987. All of these measures were largely driven by the central authority or Board of Education and were mandated without any or very little local input. As was the case in the nation, they did very little to change outcomes or increase the academic achievement of youngsters. In fact, the evidence indicates that very little changed in the area of teaching and learning (Carnegie Forum 1986).

While it is true that this period of state mandated reform served as a galvanizing force to reap further attention on reform, it clearly did not prove successful at altering the way we were doing business. The fact is that we recognized that state driven reform could only be a part of the effort to change the results of public education (Firestone, W.A., Fuhrman, S.H. and Kirst, M. W., 1989)

Another important element of the reform movement has been an attempt to change teachers. If one wants to have an impact on students, the surest way to do that is to have better teachers teaching the students. This sounds simple enough but, on second thought and on more pronounced examination, one begins to see the quality of teacher issues very differently.

First, we must all agree that teachers are a critical element in the reform agenda and, if we are to change outcomes, we must change input. However, this leads us to the question, "Are teachers doing the wrong things when they are teaching?" This question leads to [illegible] very strong reactions. Lawrence Lezotte, the school effective[illegible] guru from Michigan, says that teachers are working as hard as they can work and doing as well as they can based on their present knowledge and skills as well as the conditions under which they presently work. He says that if we are to improve their teaching, we must improve their skills and knowledge as well as the conditions under which they work. This is no mean feat since we have more than 2.6 million teachers in this country.

Of course, we could say that the way we could alter public education and its results is to train a new wave of teachers. These teachers would come to the profession from the top quarter of their graduating class, highly skilled in working with students of diverse backgrounds, very knowledgeable in their field as well as highly skilled in using advanced techniques that consider the latest research findings available in teaching and learning as well as having a very high expectation for their students. The truth is that some changes are taking place in preservice education programs, however, not nearly so dramatic as is necessary to produce the ideal teacher we just described.

Even if we were able to transform our teacher preparation institutions so that they could produce such ideal teachers we would still have two major problems. The first one is that most of the 2.6 million teachers we now have will be teaching in ten years so that new teacher preparation programs would not have an impact on the majority of teachers. The second is that most new teachers who are trained in a new way would be going into schools where old attitudes would prevail. It is much more likely that these new teachers would succumb to the approaches, attitudes, and conditions found in the majority of the existing teachers at a particular school -- the culture of the school. Thus, it is not practical to think that we can fully reform public education by creating teacher education programs that prepare a new type of teacher.

While teacher preparation programs will have an impact on a limited number of teachers over a long period of time, it seems likely that we still must reform present programs both in what they offer and how they offer it.

As such, a six-year teacher preparation program is proposed that combines a four-year bachelor's degree and a two-year master's program. In order to become a teacher in this country, a person would have to complete a six-year program. The final two years of the program, which could have broad variation from school to school, would be subsidized by the federal government and include a one-year practicum. The apprentice teacher would work in a school under the tutorage of the staff of the school. The apprentice teacher would both, be paid a stipend and receive tuition reimbursement from the federal government, and the school would receive a stipend for each teacher it accepts as an apprentice. This makes it a win-win situation for both groups and would encourage participating teachers and schools to accept apprentice teachers.

In spite of the problems described in the area of teacher preparation, there has been a great deal of attention paid to improving teacher professionalism. Two reports on improving the preparation of teachers were Tomorrow's Teachers (Holmes Group 1986), A Nation Prepared: Teacher for the 21st Century prepared by the Carnegie Forum on Education and the Economy's Task Force on Teaching as a Profession (1986). The Carnegie Foundation also has moved forward on funding and research for the development of a national teacher certification system.

This interest in teacher preparation and professionalism has naturally led to the question of teacher testing. Should teachers, as other professionals, be required to take a state examination to qualify them for a certificate that allows them to teach in that state. The discussion on teacher testing has raged in the profession with advocates for testing indicating that, while these tests do not assure quality teachers, they ascertain minimum knowledge and skills in subject areas as well as an understanding of pedagogy that would at least propose a basis for minimum competency and possible success in teaching (Madaus and Pullin 1987).

Teacher testing has proliferated in America with forty-four states having adopted some form of test requirements to be eligible for teacher certification. One central piece of criticism with respect to teacher testing has been the impact of testing on the admission of minorities into teaching. The evidence supports that minorities have not done as well on tests as majority applicants. The question of testing is still open and under discussion and scrutiny. The same questions with respect to student testing are being examined in the area of teacher testing. Will certification tests create the conditions where teacher preparation institutions in effect teach to the test? I believe the answer to this question is the same as reference to student testing. Until we have an array of testing instruments and strategies that go beyond the multiple-choice and fill-in-the-blanks variety we will find difficulty in assuring that tests are not inappropriate disqualifiers. This is one of the areas that continues to require a great deal of study and analysis before we come up with the right answer.

In view of all of this, the federal government must take some initiative in providing resources in two areas. The first is a broad analysis of assessment practices both for students and teachers with an attempt to develop instruments that deal with all of the issues I have raised as well as concentrate on assessing the new initiatives of education. The federal government must support research that will develop a new series of assessment tools. We must be able to measure the ability of teachers and students to think creatively, problem solve, negotiate solutions, work in teams, etcetera. These new assessment instruments cannot be developed haphazardly. Their development requires a commitment from the federal government to support national models. At the present time, some states and school systems are working on this issue; however, a national effort would be more cost effective. The federal government must take a leadership role in developing model teacher preparation programs that deal with these issues. This modest plan would stimulate interest in teaching and attract minorities into the profession.

It is clear that state driven reforms as described and teacher education reforms have not and will not yield the results we need to become a globally competitive nation. The question is where do we go from here in restructuring and reform if we are to do what we set out to do in the first place, which was to create public schools that can produce youngsters who can compete in the global marketplace. The fact is that only broad systemic change that touches on every part of schooling can have lasting, significant, and effective impact on results.

From a practicing superintendent's perspective, the author can identify an ongoing change model that is attempting to deal with reform from a transformational perspective. We believe that we have in Springfield embarked on a process that incorporates all of the elements necessary for systemic change to take place.

In order for real change to take place, we must understand that the place where we must look for this change is at the school level and in the classroom. All reform must move toward making the school and the classroom the unit of change. We must make schools and teachers responsible for their own destiny. This vision for our school system encompasses a change model that focuses on:

- improving student outcomes

- restructuring teaching practice
- fostering integration in all schools
- developing partnerships through collaboration and site based management.

As such, there are four major transformations that must take place. They are organizational, pedagogical, social, and attitudinal, as well as political.

Organizational

These refer to the very structure and instructional models upon which our schools are based. Our schools are presently organized around an industrial model rather than an informational model. Schools are presently organized to produce young people that are capable of working in isolation and taking direction. They are meant to produce young people who can relate to machines and not to other people. The role of the school is such that it attempts to extinguish the natural desire of people to gather, be inquisitive, and interact. Schools are organized as places where learning is a private psychological matter. The new world requires a total transformation of the organizational structure of schools.

Schools must move to become places where the organizational structure and the pedagogical models stress the importance of producing students who have the following specific skills:

- Higher thinking skills
- Be able to frame new ideas and problem solve
- Creative thinking
- Ability to conceptualize
- Be adaptable to change
- Good human relations skills
- Work in a team atmosphere
- Be able to re-learn
- Good oral communication skills
- Negotiation -- ability to build consensus, resolve conflicts

- Goal setting -- motivation, know how to get things done
- Self-assured and determined to work well
- Have many and varied work skills, including office, mechanical, and laboratory skills.
- Ability to work under pressure.
- Leadership -- ability to assume responsibility and motivate co-workers.

In order to do this, we must transform the organizational norm to one that recognizes and supports people who are able to work together and collaborate on problem identification, analysis, and solutions. In schools today, children who seek help from others are often labeled as trouble makers or even cheaters. We must organize schools in such a way that the needs of the students become the focus of the organizational structure. This means we must examine how we use time in the structure. The present practices of grade levels, scheduling, time devoted to specific subject areas, the relationship between subject areas, content coverage, length of school day and school year, and subject matter taught, must all be thoroughly examined. It is probable that the organizational structure of today's schools will be dramatically different in three years. Achieving the goal of developing problem-solving and higher order thinking skills in youth is tricky business that requires a transformation in content and pedagogy as well as in the structure of the educational enterprise.

Pedagogical

There is a growing body of evidence that indicates that present instructional delivery models cannot survive if we are to meet the needs of a twenty-first century world. It is clear that we have a growing body of knowledge about the way people learn that will strongly influence future pedagogy. These changes are not the traditional changes in methods and approaches. They are based on medical evidence that recognizes the very complex functioning of the human brain. Different people learn in different ways and it is the role of the teacher to adapt teaching techniques to learning styles. This pedagogical transformation will have a profound and lasting influence on schools and how they look in the future.

Social and Attitudinal

During the industrial society, America had a very defined set of expectations for the distribution of results. It was clear that society was controlled by a few people at the top (totally dominated by men) with most people in the middle working and taking direction from people at the top. There was a small group at the bottom who had to be taken care of by society. This group would constitute what I refer to as throw away people. The group at the bottom was in effect the excess of human capital.

These were people who our society did not need for it to be economically successful but for whom we felt a societal obligation.

As we have moved into the information society we are recognizing the need for us to change our expectation of the distribution of results. The fact is that present conditions in our country are moving us from a moral imperative to educate all to an economic imperative to educate all. American business is facing a most critical challenge in the coming century. Consider the following:

- American industry will develop 16 million new jobs by the early twenty-first century; however, it will have only 14 million people to fill these jobs.

- Of these 14 million new entrants into the workplace, a majority will be female and/or minority. This is a group that, historically, has been underprepared. A majority of these new entrants into the work force will be high risk employees. How can a country that already will have a shortage of 2 million workers cope with workers that are at risk employees and not capable of productively entering the job market? Under these circumstances, American business will not be able to survive. It becomes clear that American industry cannot afford to have at-risk workers if it is to be globally competitive.

- A majority of these 16 million new jobs will require skills far beyond those we expect of entrants into the work force today. It is estimated that 50 percent of these new jobs will require a college degree. Seventy-five percent will require at least two years of college.

While American industry today is spending between 30 and 40 billion dollars on training efforts for its employees, this investment is not enough. The schools must produce a new kind of worker for the twenty-first century who will need a new literacy and the ability to relearn and be adaptable for it is predicted that today's first graders will change jobs from four to seven times during their lifetime. Up to 50 million may need retraining in the next 10 years; 21 million new entrants plus 30 million current workers.

The truth is that America will no longer have an excess of human capital. It needs every citizen to be a productive and contributing member of society. The problem is that there is a looming mismatch between the needs of industry (the skills required of new workers) and the type of worker or student we are graduating from our schools.

American society and American schools must change their expectations of the distribution of results. People who were traditionally not expected to succeed must now succeed if our economy is to survive. This requires a complete social and attitudinal transformation on the part of our society and more specifically our teachers. The challenge has now become not teaching children to the best of their potential but teaching students to the best of our potential. The new paradigm indicates that it is what we do in the schools in response to how the children come to school that makes the difference and not how they come to school. This transformation is possibly the most challenging and the most difficult for the American public school to make.

Political

This area of transformation has several parts and includes political change within the school construct as well as in government and society in general. First, it is important that we recognize that we live in a society that has had as its underpinning a strong middle class. This middle class as of late has not been replenishing itself. An analysis of our national birth rate indicates that the middle class is having about one and one-half babies per marriage. This means that the natural replenishment of the middle class is not taking place. By comparison, the birthrate for poor people is exploding. The growing sector in this country is the children of the poor.

The political question here surrounds the will of this country to educate those that it has traditionally ignored. Will American society understand the political and economic repercussions and implications of not educating its poor? Will American society support public education in urban centers when the people being educated do not resemble both in class and color the people controlling the economics of those urban centers?

The additional fundamental issue of equity and excellence must also be addressed within the political context. At the present, where you are born will to a great extent determine the quality of your education. There are communities in this country that spend $1,200 a year per child while others spend as high as $14,000 per child. While we understand the issue is not money alone, how could anyone accept that there is not an inherent political inequality in this funding approach.

A political transformation is required at the local and federal level in the area of funding public education. We cannot continue to run away from this reality. This is the political issue of our times that must be confronted very soon in this country.

An additional political transformation that must take place revolves around what we teach our children and how we measure what we teach. The how we teach them was dealt with in the pedagogical and organizational transformation. We must come to some political agreement on what we expect our children to know and how we will measure what they know. These two areas demand broad national attention and must be resolved politically.

The final political transformation requires American public education and its governance to remain at the local level. All attempts to nationalize education are filled with danger; however, America must develop a federal funding process that is supportive of an equal education for all. This is one of the major areas of political transformation that must take place during the 1990s.

We as a nation must develop a plan to improve education that includes financial support to deal with all of the issues that face our children. We must combine the appropriate distribution of money with adequate accountability so that money would not be wasted as is the case in so many federal programs.

As I have already indicated, the single most critical issue in education today is one of equity. Does every child born in America have equal access to an effective and appropriate education? Our present system is such that if you are born poor, you will more than likely receive an inferior education. The difference between what is spent on poor children and what is spent on middle class children, as previously indicated, is immense. Moreover, the research is clearly in support of the implementation of early childhood programs that provide a firm foundation for continued development and academic achievement. Why not begin all schooling at age 4 and continue for 13 years? This change in age would not increase the number of years of K-12 education, but would provide education during those important formative years, and would allow students to end at age 17. Then, they can continue learning as an apprentice at a job or continue a postsecondary education. All that we know about the changes in society and the workplace indicate that the worker of tomorrow must be capable in many skill areas and must have higher thinking ability. Beginning earlier and providing a continuum of educational opportunities will go a long way in addressing these new challenges.

We are at the crossroads of choosing to pay adequately for the education of all children regardless of where they live, the color of their skin, or the language they speak.

The federal government must play a more intensive role in the funding of American public education. The link between our economic survival as a nation and education has been clearly defined. The question is more how America can raise funds for accomplishing this task. We must institute a tax program that specifically raises funds for education. The author proposes a U.S. Mail Education Surcharge. Why not a 15 cent education surcharge on every piece of mail with a higher scale for pieces of mail that cost over one dollar? This education tax would affect every individual and every business in our nation. An equitable distribution plan for this money would also be easy to devise.

These transformations can take place in America if we understand and accept the following precepts:

1. Money is not the answer, but without money we cannot do the job.
2. Children do not come to school the same way; however, it is our response to how they come that makes the difference.
3. Some children cost more to educate than others. It is in our best interest to educate them all.
4. The present system of funding public education is inequitable and must be changed. Where you are born to a great extent determines how much will be spent to educate you.
5. The classroom and school is the unit of change and, as such, local governance must be promoted, encouraged, and maintained.
6. The present model of education must be adjusted so that first time quality becomes the norm and not remediation as is presently the case. Thus, schools must change their focus. Education or Schooling should begin at 4 years old for all youngsters. This can be done without spending additional money. All we would have to do is rearrange our present curriculum and keep kids in school for 13 years; just begin one year earlier. This would take several years to implement.
7. The relationship between the school, the home, and the community must be understood and internalized. Schools need the community and the community needs the schools. They cannot exist independent of each other.

8. We must realign our goals with our curriculum. What is it that students really need to know for the twenty-first century? It is insane and silly to teach well what these students cannot use. Every community must ask itself what do we want our children to know? What will we accept as evidence that they have learned? How can we measure what they have learned? Multidimensional assessments must be developed to accomplish this task.

9. Our classrooms and the way they look and are organized must change dramatically. We know enough to do this right now. The research on how children learn is exploding before our eyes; yet we have not implemented one-tenth of what we know about learning and teaching.

10. We have not focused on technology as the key to the future. We are not using even one-tenth of the power of technology. We must move from the chalkboard to the electronic whiteboard. We must integrate learning areas around the technology that exists.

11. We must learn the principle of organized abandonment. Abandon the things that have not worked for a long time such as age grade grouping, retention, tracking, standardized tests, the Carnegie unit as a process and not a product unit; we must abandon our present system of scheduling, particularly at the high school level. We must abandon specific student to teacher ratio and let teachers decide what is necessary, appropriate, and effective.

12. We must transform our schools from places where people are told what to do, to places where students, parents, teachers, and administrators identify the issues and provide the solutions. These constituencies must be able to exercise control over their own destiny. With this control and power will come increased accountability. As we provide the staff with this empowerment, they will be able to greatly influence learning. This should naturally lead them to commanding higher salaries and status.

13. We must use choice as a school reform methodology with great care. It must be crafted so that it does not lead to additional inequities for a segment of our population or as a divider of the haves and have nots.

14. We need massive teacher training programs that will help teachers understand the new ways available to educate their youngsters. This must be done at the school level and planned and developed by teachers.

15. We need additional time in the school day where teachers can

plan together around the issues that confront them. Schools must become units of change where teachers see the interdependence of what they teach and how they work and support each other.

16. We need to have everyone in America understand the seriousness of our work and the interdependence of the quality of life in our community and the quality of our schools. We need as a nation to understand the relationship between quality education and the salvation of our democracy.

In Springfield, Massachusetts, a city of 165,000 people with a school population of 25,000 students, 30 percent which are Hispanic, 30 percent which are Afro-American and 40 percent which are white, we began a restructuring effort in September 1989 at all forty schools, which centered around the four transformational efforts we have previously described.

When the author came to Springfield in September of 1989, he was given a charge by the School Committee to bring broad, comprehensive, and systemic change to the school system. They had been struck by his comment during the interview process that said if you want to keep getting what you have been getting, keep doing what you have been doing. If you want new results, you have to dramatically change what you are doing. The system was ready for change and that change process was detailed in a report called Blueprint for Excellence presented to the community in November 1989. The report was a blueprint for change that would be adjusted with the broad input of all the constituencies in the community.

The change process had as its main focus the improvement of the schools through collaboration and cooperation. In order to focus the attention of the community on what had to be done, four talking papers were prepared and four task forces were created with representatives from every segment of the community. The four task forces were:

1. Central Office Reorganization.
2. Restructuring of Grades
3. Curriculum For the Twenty-First Century
4. Effective Schools Research and Implementation

The four task force reports formed the cornerstone for the work accomplished during the first and second year of this reform effort.

The Central Office reorganization led to a more streamlined Central Office. Since there was to be a massive shift in authority to the schools, the responsibilities of the Central Office would change. The Central Office took on a new role. It moved from the role of director to assistor. Rather than telling people what to do, we became enablers, resource providers, or facilitators. This was and continues to be one of the most difficult paradigm shifts in the system.

The second task force came to its conclusions very swiftly in that everyone knew that the K-4, 5-6, 7-9, and 10-12 organizational structure was antiquated and not working. The task force recommended a conversion to a K-5, 6-8, two K-8 schools, and 9-12 system. In order to do this and also improve our integration efforts, which were out of compliance, we developed a Schools of Choice Plan that incorporated elements from controlled choice plans that had been implemented in several cities across the country. The changes, therefore, included a new grade structure for the entire system, a controlled choice plan where each school had developed a very specific theme that made distinct, the conversion from a junior high school philosophy and approach to a middle school concept, as well as the conversion of all 10 to 12 schools to a grade 9-12 high school system. This required the closing of a junior high school and its conversion to a 9th grade annex for a high school. All of the students in the system were given a choice in the selection of four possible schools. Eighty-four percent of the parents received their first choice and dramatic improvement was made in the area of racial balance.

A volunteer transfer plan was developed with the Teachers' Union, and more than 450 teachers were transferred to the schools of their choice. In addition, a special agreement with the Supervisors' Union led to the movement of 10 Central Office administrators to school based supervisory positions.

The third task force Curriculum for the Twenty-First Century reaffirmed all of the recommendations in the talking paper which centered around realigning the curriculum so that specific goals and objectives were very defined for teachers. It included the implementation of a technology based model of assessment that would be developed by teachers. The plan outlined the expectations of a twenty-first century curriculum which included a detailed process for periodic review in the area of comprehensiveness, authenticity, and quality. It also incorporates a broad school centered staff development program as part of the process.

The emphasis of this task force was to create a curriculum process that emphasizes students' ability to reproduce and use knowledge. It also stresses a curriculum for all of the children and not for one that is tailored to a chosen few.

The Effective Schools Research Task Force reaffirmed the intent of the talking paper and went several steps further. It recognized the need to incorporate into the very culture of the system the finding of effective schools research. It led to Lawrence Lezotte and

James Comer becoming major consultants to the school system. It also stressed the importance of the continued decentralization of the school system through school centered decision making. All 40 schools now have site-based teams that have begun to be responsible for the operation of the school. During contract negotiations with the teachers for the 1991-92 school year, a letter was signed that formally introduced site-based management and teacher empowerment into the contract.

In addition to this, the union, the central administration, supervisors and the business community have begun negotiations around the introduction of a total quality management program in the system.

The superintendent and the entire staff of the Springfield Public Schools were steadfast in our goal to provide an equitable education for all. We could not stand still to await a better day. We decided, in spite of drastic budget cuts, to identify the major issues confronting our school system. We recognized our growing student population and the need to fulfill the promise that had been made to the community of a K-8 magnet school; therefore, we pursued plans to build this new school. The plans were formally approved by the state and we broke ground in March for a school that will open in 1991 for 1,000 students. We implemented the Schools of Choice Plan that involved the community in an unprecedented way. At a Schools Fair which opened a three week application period for all students in kindergarten through grade 9, more than 10,000 parents came, reviewed the school booths, and spoke with staff. During the ensuing weeks, thousands of parents visited the schools they were considering. A new era in public education in Springfield had begun. These parents who sought the best school for their child will also continue to be involved at the school to ensure that the promise is realized.

In September 1991, the Springfield Public Schools opened with all grades restructured. Parents could choose a kindergarten through grade 5 school within their educational zone or a kindergarten through grade 8 school in a city-wide magnet offering. Middle schools, grades 6-8, were made accessible to all students city-wide as well as the high schools, grades 9-12.

But we all know that this was just the beginning. The framework for excellence in education was set. The results, however, in student achievement are affected by more than grade structure. We have explored alternative solutions to ensure teaching for learning as well as equity and excellence for all children. We have the will to make each school an improving school. Though we have just begun, our journey is clear and direct to making every school in Springfield work for all its children.

In Phase I, we set the framework for school improvement with the Blueprint for Excellence, which identified areas of immediate concern and long-range planning -- all of which included all of the constituencies in a policy of inclusion.

The policy of inclusion included the task forces that were previously described; however, at the same time, we developed very specific initiatives to involve the community in our schools. We believe that schools cannot exist in isolation of the community. A community cannot have an effective quality of life without effective schools to support that quality of life.

As such, we developed four major initiatives that would stress the involvement of the broad community. They were Parental Involvement, the Conference for Children, the Business Education Agreement, and the Religious Community Initiative. In the first initiative, we sat with groups of parents and created a parent involvement policy that was truly revolutionary. It created the Springfield Parent Advisory Network (SPAN) which would be an organization that represented all of the parents in Springfield. The policy that was adopted by the school committee created a working parents organization in every school as a requirement of the system. In addition, the parents have been provided professional organizing assistance paid for by the school system. This has created an independent organization that acts as an advocate for children and families.

The Conference for Children was an initiative that convened more than 300 public and private service providers. The intent of the conference was to develop a process or institution in the city that would become responsible for making the city a child-centered city. These 300 agencies and individuals signed a document that created the Alliance for Youth in the city. A board of directors with representatives from the highest level sits on this board and gives direction to the Alliance for Youth. The Alliance has already developed several major initiatives for the children of the city including a conflict resolution-violence prevention program for students in the middle schools. The Alliance has served a major role in having private and public agencies provide direct services to children in the area of drug prevention, AIDS education, mental health services as well as child abuse prevention services.

The Business-Education Agreement was developed by the local Chamber of Commerce and the school system to address the issues confronting the schools and the business community. It clearly enumerates the role of business in improving the schools as well as the responsibilities and accountability of the school system. We have more than 60 companies participating in a variety of relationships with the schools.

The final initiative is possibly the most unique and thought to be the first of its kind in America. It convened over 100 religious leaders at a conference where they signed an agreement relative to how they would collaborate with the Springfield Public Schools. It outlined specific steps that the religious community would take to support the public schools. We have agreed as a major undertaking to support together the issue of social justice for all people. We are currently planning specific programs to implement this goal.

In Phase II, we continued dialogue with all the constituencies, addressed program design, redefined responsibilities, trained for new roles and teaching techniques, and implemented a Schools of Choice Plan.

Phase III, during 1991-92, will involve the implementation of solutions, continued training for all constituencies to prepare them for new roles of involvement, and the establishment of task forces in four critical planning areas -- early childhood, high schools, technology, and retention and tracking.

We have the capacity and the will to make Springfield the first city in the nation with an effective school system. We recognize the changing societal demands and influences on our students. We know what must change within the schools. New interventions and strategies on how to teach as well as renewed commitment and energy are focused on school improvement.

My challenge is for every American to take risks, to act boldly, to say our children must be saved. It is this philosophy that has been applied in the School Improvement Plan for the Springfield Public Schools. It would have been too easy to say we cannot try to better our educational program as we faced massive budget cuts in many areas; it would have been too easy to say that the state of our schools and our society is too complex for immediate positive results.

For those who recognize the need for change to meet the inevitable challenges of future life but wish to slow down the process, I say that world events and local implications are on an accelerated time piece.

Beyond the evident changes, we see reversals:

- countries that limited movement of their citizens now advocate freedom;
- places where capitalism was a bad word now embrace the basic tenets to address their people's deprivation;

- core curriculum changes that include and embrace non-western cultures and works;
- career plans that suited a life time must now incorporate multiple skills and directions;
- traditional family structures are being challenged by alternative structures;
- limited expectations for females relegated to an ideal view of the home has changed to allow equal access to careers -- not due to equity but to necessity both in the home and the workplace;
- isolation of the races and mobility for limited groups no longer works in a pluralistic society that requires all for economic and social success.

In 1970, John Holt wrote in What Do I Do Monday? -- "Every day's headlines show more clearly that the old ways, the "tried and true" ways, are simply and quite spectacularly not working. No point in arguing about who's to blame. The time has come to do something very different. The way to begin is -- to begin." Two decades is long enough to wait to begin. In Springfield, we cannot continue to accept a 40 percent dropout rate (60 percent among Hispanics); we cannot continue to blame others for the lack of individual success without addressing that which we can control; we cannot assume that the curriculum and methods of the past will serve us well in the present since those of us here are the survivors of a system that did not attempt to educate all children. We cannot postpone what is morally right.

I consider these accomplishments to be outstanding feats for such a short period. Of course there is a down side to this as there is to every story. The systemic changes necessary to institutionalize all of this has not taken place yet. There is a reluctance to give up the old and more importantly to relinquish power. There is a hesitancy at every juncture to agree to a process that will allow for multiple inputs and shared decision making if it leads to the loss of power.

What is required is incremental change and frequent small successes (measures of growth) that one can point to as the basis for further movement. Our role is to make the system work for the people it serves and not the people who run it. However, public institutions or systems, do not and will not reform themselves. They need coaxing, coercing, and reasons to change. They need outside intervention. My sense is that this bold experiment can work. Not enough has been done yet in Springfield to merit distinction -- but we do merit watching.

References

Carnegie Forum on Education and the Economy (1986) A Nation Prepared: Teachers for the Twenty-first Century (New York: Carnegie Corporation).

Chubb, J.E. and Moe, T. (1990) Politics, Markets and American Schools (Washington, DC: Brookings Institute).

Darling-Hammond, L. and Berry, B. (1988) Evolution of Teacher Policy. Report of the Center for Policy Research in Education, Eagleton Institute of Politics at Rutgers University and The Rand Corporation, Washington, DC.

Elmore, R.F. and associates (1990) Restructuring Schools: The Generation of Education Reform (San Francisco, CA:Jossey-Bass).

Firestone, W.A., Fuhrman, S.H. and Kirst, M.W. (1989) The Progress of Reform: An Appraisal of State Education Initiatives (New Brunswick, NJ: Rutgers University, Center for Policy Research in Education).

Fuhrman, S.H. and Malen, B. (1991) The Politics of Curriculum and Testing, New York: Falmer Press.

Holmes Group (1990) Tomorrow's Schools: A report from the Holmes Group (East Lansing, MI: The Holmes Group).

Murphy, J. (1990) The Educational Reform Movement of the 1980's. California: McCutchan Corporation.

National Commission on Excellence in Education (1983). A nation at risk. The imperative of educational reform. Washington, DC: National Commission on Excellence.

Will the LEP Train Reach Its Destination?

Designing an IHE Teacher Training Program for Specific LEP Student Instructional Needs

John E. Steffens
University of Oklahoma, Norman

As we Americans enter the last decade approaching the year 2000, the various agencies related to Education, Labor, Health, Human Services, etc. are challenged by the enormity of the tasks ahead to achieve equitable quality of life for all citizens of America. A concomitant general sense of commitment to accomplish these goals is being evidenced by citizenry and futuristic demographers alike. A host of recent publications emphasize a national commitment toward changing positively access and effect of societal programs to serve the whole of the American population.

This paper is intended to:

- Excite you to want to participate in building a new paradigm in teacher training for teachers of all LEP students;
- Provide background and a preliminary knowledge base to substantiate a call for action;
- Relate the need for paying attention to LEP students in educational reform and restructuring activities, particularly the AMERICA 2000 strategies; and,
- Describe some steps that need to be taken now to accomplish the tasks outlined.

Building A New Paradigm

A new way must be defined to look at what all teachers need to know and be able to do when LEP students are assigned to their classes. Operational programs set into place in all teacher training institutions should build upon what has been proved successful and demonstrated to work in teaching LEP students. Administrative strategies should provide total administrative and community support to assure the educational advancement and required services for LEP students.

The need to engage professorial and administrative education personnel nationwide in the dialogue to develop this new paradigm is crucial. The ability to bring together a common knowledge base for teaching LEP children is vitally necessary. The strategies for providing equal education opportunities for all LEP students, wherever they choose to attend school, is our national responsibility. By working together, we can identify what works in teaching and supporting LEP students. We can build this knowledge base into the teacher training curricula of the one thousand or more teacher training institutions of the nation. Every emerging teacher will then have a foundation of what to do when a LEP student is assigned to the classroom.

What should MATH teachers know about teaching math to LEP students?

What should GEOGRAPHY teachers know about teaching geography to LEP students?

What should HISTORY teachers know about teaching history to LEP students?

What should SCIENCE teachers know about teaching science to LEP students?

What should ENGLISH teachers know about teaching English to LEP students?

Education Secretary Lamar Alexander describes four trains (AMERICA 2000, p. 12) running on parallel tracks, each headed toward educational excellence! The four trains represent the four parts of the AMERICA 2000 strategy.

We would expect that each of the four trains should be in excellent mechanical condition to arrive at its destination by the year 2000. If a critical part of the train is defective, however, the whole train might be delayed. I suggest to you that each of the trains has a defective wheel that needs to be repaired. The defective wheel is supported by a defective undercarriage. The undercarriage supports a car full of LEP students. The defective undercarriage represents the teaching and education services for individual LEP students where no bilingual classes are offered. The defective wheel is the teacher training of all teachers who will have one or more LEP students in their classrooms. The car represents the education curriculum in each of the five core subjects to support individual LEP students. The passengers in the car are LEP students.

The train cannot arrive at its destination without its precious cargo. The expectations are that if the train does not arrive, the AMERICA 2000 strategy is a failure -- by its own definitions and standards! How to fix the defective wheel and undercarriage can be found in quality sciences and voluntary standards, using a process to establish and use a new paradigm. Attitudinal shifts must occur, such as:

FROM	TO
• Add-on nuisance	• Essential to my job
• They don't count.	• I can't succeed if they don't.

From a world perspective, many in education were startled by findings reported by the International Association for the Evaluation of Educational Achievement. According to their reports:[1]

> ...Assessments of 20 school systems around the world rank American eighth graders 10th in arithmetic, 12th in algebra, and 16th in geometry. Even America's top students fare poorly in international comparisons: among the top 1 percent of high school seniors, American students ranked last.[2]
>
> Achievement in science is no better. Among 10-year-olds in 15 countries, Americans rank eighth. Among 14-year-olds in 17 countries, Americans tie with children in Singapore and Thailand for 14th place. Among advanced science students in 12 nations, Americans are 11th in chemistry, 9th in physics, and last in biology.[3]

These statistics are only a sample supporting the general conclusion that much must be accomplished to improve education if we are to meet the educational necessities for **all** Americans.

The changing demographics of the United States is apparent to even the casual observer. "Language minority children make up a growing proportion of U.S. youngsters. It is estimated that the number of such children aged birth to [four] years rose from 1.8 million in 1976 to 2.6 million in 1990 (Soto, 1991). The number of children with limited English proficiency is expected to continue to increase."[4]

Many of these children, from various ethnolinguistic backgrounds are identifiable as limited English proficient (LEP) students. These LEP students are not concentrated in any one location and in any one environment. A portion of these students reside where significant numbers of students are of similar ethnolinguistic backgrounds. Others reside in communities where significant numbers of LEP students come from a variety of different ethnolinguistic backgrounds. Another category of LEP students are those who reside in small groups scattered across this country, sparsely distributed in communities and schools so that sometimes only one, or on occasion a few or even several LEP students of different ethnolinguistic backgrounds might be found in the classrooms of this country.

The problem of providing adequate and equal service for these students is even more complex. The shortages in the supply of bilingual teachers to fill bilingual classrooms described in the first two scenarios of the previous paragraph have been documented in the literature. For those two categories, i.e., large groups of same language and large groups of different languages LEP students, the needs have been identified, must be met, and discussions and descriptions of appropriate teacher training programs to meet the needs have been developed. This has all received previous attention in both literature and program implementations, and certainly deserves continued consideration, both because of current need and also because of increasing demand and future growth projections.

This paper, however, will focus on that portion of LEP students fitting the description of residing in those communities where students with various non-English speaking ethnolinguistic backgrounds are sparsely distributed. As in other parts of the American culture, this specific LEP population is rapidly growing. The majority of classrooms in America have from one to several of such students in them. With 110,000 schools in this country, the specific LEP student population that has not been adequately served and needs to be served represents a significant number of children. Though not as visible because they are more sparsely distributed, the reality of cultural and education shock and/or adjustment/accommodation is just as significant -- and sometimes may be even more so -- for a LEP child in this less concentrated environment rather than the more highly concentrated LEP environment presently receiving the most study, attention, and services.

> In general, children whose first language, or whose families' first language, is not English score lower than their English-proficient peers on standardized reading and math tests.[5] By third grade, children whose families often or always speak a language other than English at home may be more than a year behind their peers in reading proficiency.[6]

If by the third grade, LEP children are one year or more behind their peers in reading proficiency, it follows that these students are very "high risk" students for dropping out of school in the future as well as high risk for a full assortment of other risk behaviors (pregnancy, alcohol/drug abuse, etc.)

Robert Milk[7] concisely summarizes recent literature and its application to the task before us in preparing teachers to appropriately meet the educational needs of LEP students. He suggests that:

> One clear theme that emerges from contemporary discussions on preparation of teachers for mainstream education is that programs need to achieve greater integration of theory and practice.

> This [concept] is [supported] in the language teaching literature (Alatis, Stern and Strevens, 1983).[8]

Methods courses must stress the interrelationship of theory and practice. In addition, experiential activities must provide hands-on field experience for the effective preparation of a teacher (Mellgren, Walker and Lange, 1988; Celce-Murcia, 1983; McGroarty and Galvan, 1985; Clark and Milk, 1984).[9]

Milk cites that it is important to develop a research perspective in future teachers that will encourage them to be curious, to ask questions as to what is happening in the learning environment, to observe closely, and to develop a heightened awareness about what is occurring. He also refers to the need for a balanced amount of intuition.

Teachers must experience preparation which provides interrelated knowledge and experiences drawing from linguistics, psychology, sociology, and culture (Politzer, 1978:14). At many institutions this may represent a need to collaborate across the disciplines and/or departmental lines of education, foreign language, linguistics, English or even more (Milk, 1985). There is significant support for integrating the areas of bilingual education, ESL, and foreign language in the preparation of teachers. McKeon (1985) found a significant overlap in teacher education standards in these areas and also found common research themes across the three areas. Collier suggests that course work to prepare ESL and bilingual teachers is similar in many ways and "bilingual and ESL staff can benefit most from an integrated approach to training."[10]

Educational Reform and Restructuring

The focus of this paper is to address the need for the training of teachers who will have responsibility for teaching students from specific LEP populations sparsely scattered throughout American classrooms where numbers are not concentrated enough to support bilingual class structures and teachers as such. The scenario presented will describe how teacher training should take place to affect the educational experience across the multitude of school communities and classrooms in the United States. Positive education outcomes must be a reality for the LEP children who are distributed sparsely throughout the schools and classrooms of America. To do anything less is to fail.

The author of this paper asserts that such preparation of teachers must happen within the context of what is known to work in the areas of bilingual education, ESL, whole language learning, etc. involving collaborative contributions of linguistics, psychology, sociology, culture, organizational management, social work, and academic content.

If America is to achieve the "year 2000" goals in education -- health, employment, youth -- we, in this country, will be required to approach these areas in new and different ways. It will require that teacher trainers experience what has become known as a "paradigm shift." The teaching profession will be required to envision the entire education and social scenario in a new kind of way. Then, within this new envisioning, develop missions, goals, standards, objectives, strategic plans, curriculums, activities, and assessments. This process must include what we now know regarding bilingual/bicultural education, identify the areas where standards must be set and met, and engage in the dynamic process to ensure the accomplishment of the process and tasks.

Within the context of this paper, the author describes an approach which will contribute to the education of and "make a difference" in the lives of LEP children in classrooms where few, if any other LEP students, are present.

1. If the assumption is true that in the majority of classrooms in this country there are one or more LEP children and there are 110,000 schools, then the population of LEP children totals tens of thousands of students -- and the number is growing rapidly.

2. If the assumption is true that we cannot supply enough bilingual teachers even for existing bilingual classrooms, then we certainly have not been able to supply adequately prepared bilingual teachers for these classrooms with smaller numbers of LEP students, either.

3. If the assumption is true that we must develop the local educational environment to adequately serve LEP students throughout America so that the educational achievement of all LEP students is enhanced and not inhibited, then we must plan, design, and implement an education process for developing the programs whose foundations are rooted in the "known," but whose delivery is structured under a new paradigm.

As we contemplate what the response to this third assumption might be, some questions arise. What might such a new design for a teacher training paradigm look like? How could American education possibly meet such a challenge? To succeed we must rely on what is known and apply it in a new kind of way. We must utilize the contributions of education, management theory and practice, social change and social systems knowledge, and sociology and psychology to establish and use a new knowledge base for preparing teachers of LEP students.

One dimension of the new paradigm that must be addressed relates to sheer numbers of students. If most classrooms in America either have or will have one or more LEP students, it follows that all teachers must be prepared to ensure that the learning environment and education practices enhance educational achievement for LEP students and ensure that no inhibition of education occurs.

For this to be accomplished, we cannot rely only on receiving the services of the various centers and in-service and preservice bilingual training programs that currently exist. These programs have been repeatedly proven useful and successful and the magnitude of need for these programs continues to grow as the LEP populations multiply. Therefore, another supplemental approach that holds promise is to focus on the foundation block of teacher training. There are approximately 1,000 colleges of education throughout the United States. Only a fraction of these have bilingual education preparation programs. If we are to change the paradigm of American education for these LEP children, we must implement a systemic approach which will facilitate change for *all* educators. We must consider both the organizational management dimensions as well as the content or input necessary to ensure the transformation toward a facilitative educational experience for LEP children. The results will most certainly assure an equitable educational outcome for LEP students.

First of all, let us address the managerial side. The majority of teachers in America are produced by the many regional teacher training institutions throughout the states of this country. The foundational structure for the changes necessary in the academic approaches and the content areas in these teacher training programs involve the development of a new paradigm for the content methodology and procedures related to the educational sequences in teacher preparation, educational leadership and administrative preparation programs. The paradigm shift for management of teacher education must include the comprehensive content, that is, the total outcome of the teacher training enterprises. Attention must be paid to what *every* teacher needs to know and be able to do in working with LEP students. Professors and the higher education community responsible for these training sequences must review, revise, and implement the necessary changes to ensure that all educators with whom they have contact become prepared to respond according to the new paradigm, and do so as it is being defined and established.

Regular classroom teachers working with one or more LEP students must be informed and practiced in the art and science of teaching LEP students. Thonis (1991) has identified characteristics that teachers who work with LEP students should possess.[11]

- an awareness of cultural differences
- a recognition of language diversity

- a knowledge of second language acquisition theory
- an understanding of the students' realities
- a sensitivity to the values of families
- a knowledge of the history and heritage of the group
- a recognition of strengths and potential of all students
- a willingness to modify and adapt instruction as needed
- a solid grasp of curriculum imperatives for students learning in a second language.

As we move toward the paradigm shift, we might ask ourselves how the shift could be accomplished. Managerially, this shift might be addressed by a series of summer institutes for IHE faculty and administrators designed to increase faculty knowledge and perception of necessary theory and practice involving LEP students. Upon acquisition of this input, faculty would revise methodology courses to include necessary content and practice. These faculty would then return to their institutions with a four-point charge:

1. Implement the curricular changes into the scope and sequence of teacher preparation at their IHE.
2. In-service their own faculty in these curricular changes.
3. In-service teachers in schools in the local service area regularly served by the local IHE.
4. Participate in an ongoing national dialogue to define the new paradigm and adjust as a national agreement emerges.

What should be included in this new program? What are some dimensions which must be addressed?

The U.S. Department of Education (USDE) guide for implementing the first national goal cites children from families where English is not spoken require schools and communities to develop new ways of educating children and securing the support of their families.[12] This report further suggests that the involvement of parents is critical to the development of young children and their educational success.[13] And that while proficiency in more than one language is a lifelong resource, children whose English proficiency is limited need special assistance as they prepare for school success.[14] And that developmentally appropriate, culturally sensitive programs should be available.

882

Decades of research on successful or effective schools identify several common characteristics. Effective schools have high expectations for students and teachers. They set rigorous academic standards, maintain order and discipline, require homework, and encourage parental support and cooperation.[15] They have strong leadership from a principal; a stable staff of competent and enthusiastic teachers; a curriculum that is integrated across grade levels and that accommodates the variety of learning styles and cultural backgrounds of their students; and opportunities for parents to participate in their children's education. Underlying all of these elements is a set of clear and broadly accepted educational goals -- a vision or mission to which all members of the school community are committed.[16]

Research on effective schools also stresses the importance of school climate -- the physical and social environment in which education takes place. At a minimum, school climate refers to physically safe and personally supportive schools and classrooms and mutual respect between students and educators.[17] More broadly, a positive school climate refers to classroom and learning environments that make it possible for students and teachers to work toward the common goals or shared educational mission of the school. It is also characterized by active involvement by parents and teachers in important school decisions.[18]

Numerous recent reports support the concept that education is a social phenomenon involving the whole community. However, in the past, schools have tended to regard themselves and be regarded by law and social policy as "isolated, disconnected segments of our social and economic lives."[19] Society has "put a disproportionate faith in the impact of schools working alone" to solve educational problems.[20] Yet, a review of the education literature suggests that educators, working alone, cannot possibly solve the multi-faceted and complex societal challenges. It is becoming increasingly recognized that "in order to effectively meet these challenges, the entire community must be involved: parents, schools, students, law enforcement authorities, religious groups, social service agencies, and the media. This broad-based approach -- one that has achieved successful results related to our nation's recent school improvement and educational excellence movements -- involves bringing all available human and material resources to bear on the situation at hand.[21]

The recent proliferation of educational activities throughout the United States is viewed as both an expression of public commitment to action and representative of a vast resource of talent, commitment, and ideas. Yet, it should be noted that:

> When educational institutions and agencies undertake collaborative efforts in education, an initial tendency is to enter into discussions about how one agency can help the other(s). The pre-

> dominant notion is that individuals in one setting are more skilled, possess more accurate insights, are better equipped to bring about a desired improvement than those in the other settings. The less the collaborators have worked together in the past, the more this attitude appears to prevail in the minds of both the people in the schools and those in other agencies.
>
> As a result, a work on rather than a work with posture underlies many joint efforts....[22]

The organizational development that a local community must undergo in responding to the current educational crisis and the wide variety of skills needed to plan and implement initiatives require maximum commitment and participation. Many school district personnel already possess much of the knowledge and many of the human resource skills needed to create and operationalize an effective plan. However, a new paradigm that incorporates the latest knowledge in school effectiveness has not been developed and accepted by many schools and teacher training institutions.

It is important to note that some school personnel are involved in the surrounding community activities and organizations. These "boundary spanners" have one foot in the school system and the other in the infrastructure of the surrounding community. As such, they are able to identify individuals, organizations, and social groups in the community.[23] From this pool of potential resources can be assembled individuals who will be invaluable in identifying and mobilizing other human and material resources in the community. Through their efforts, a collaborated vision of a new reality can emerge, a new paradigm for school effectiveness can become operational.

Educators have found that, by involving people right from the beginning, their communities are more likely to come together and work cooperatively with the schools in achieving the goals they have formulated together.[24] People who are involved from the start are committed to a shared vision of what a school should be, and work to make that vision reality.

Let us now address some of the general content areas that teacher trainers in university teacher education programs should provide as a framework for training public school teachers in the skills and knowledge that will prepare them to address the needs of multicultural student populations. Specifically, these areas provide the necessary information and resources to introduce multicultural education training into the teacher education curricula. These items should most appropriately be inserted into the teacher training curricula, rather than segmented onto it. Every American teacher should know about and be able to do certain activities to support the schooling growth of LEP students. The preliminary list provided below was developed by Dr. Ravi Sheorey, assistant director of the Service Area Eight Bilingual Education Multifunctional Resource Center at the University of Oklahoma, from a variety of sources, to initiate thought and reflection.

1. Introduction

Major terms and concepts in multicultural education

Ethnolinguistic diversity and American public schools: A demographic profile and projection for the 1990s

Language diversity and public school education: The needs of limited English proficient (LEP) students

Educational equity, cultural pluralism, and multicultural education

The need for a multicultural education component in teacher education programs

11. An Historical Overview of Multicultural Education

Multicultural education in non-U.S. Western industrialized countries

Multicultural education in the U.S. in the 19th and 20th centuries

Multicultural education in the "global village": The case of the U.S.

111. Multicultural Education and Related Issues

Language policy in the United States: past and present

The relationship of language and culture

Teaching and learning native and second languages

Native language maintenance: help or hindrance to education?

The role of language and culture in cognitive development and selfconcept development

IV. Bilingual Schooling and Multicultural Education

The rationale for bilingual/multicultural education

Bilingual education programs in the United States: Federal laws and their implementation in schools

Recent trends in bilingual programs and practices

Major research findings about the effectiveness of bilingual education in American schools

V. Assessment Issues in Bilingual/Multicultural Education

Language proficiency, bilingualism, and academic development

Referral, assessment, and placement of language-minority students in public schools

The construct of language proficiency: communicative versus academic language proficiency

Developing "culture-fair" assessment procedures

Testing LEP students in English and the native language

VI. Multicultural Education and Special Education

The construct of learning disability and the LEP student

The measurement of learning disabilities in multicultural education

Patterns of special education placement of culturally diverse students

VII. Developing a Multicultural Curriculum in Teacher Education

The rationale for curricular adjustment in teacher education programs

Multiculturalism in the curricula related to the teaching of math and science, social studies and language arts

Introducing cross-cultural variables in teacher education courses

886

Infusing multiculturism in the field experiences of prospective teachers

VIII. Competencies for Prospective Teachers in Multicultural Education

Personality attributes

Affective skills

Pedagogical skills

Cross-cultural field experiences

IX. Teaching Strategies for Multicultural Education

Self-assessment of multicultural education skills

Values, perceptions, and assumptions in various ethnic groups

Cross-cultural communication: verbal and non-verbal

"Hands-on" training methodologies: Simulations, role-playing, critical incident/case study approaches, decision-making in a cross-cultural setting, etc.

X. Evaluation of Multicultural Education Component in Teacher Training

Entities to be evaluated: knowledge, perceptions, attitudes, skills, and patterns of behavior

Techniques of evaluation: paper and pencil exercises, critical incidents, self-analysis reports, etc.

Measurement of changes in attitudes and perception at the beginning and end of program

Effort is required to determine appropriate administrative principles and practices, to synthesize the components of a school into an effective organization and to meet these challenges. The effort for defining and achieving quality in the process is a continuing one.

Besides the increasing complexity of the teaching profession, it is also becoming increasingly more challenging to determine administrative principles and practices which effectively tie the behavioral variables of an organization into harmonious and productive units.

Guba indicates that the unique task of the administrator can be understood as that of mediating between the behavior eliciting forces of organization needs and individual needs so as to produce behavior which is organizationally useful as well as individually satisfying. Action leading to such behavior on the part of individual members is the highest expression of the administrator's art.[25] Likert reinforces this view by insisting that it is essential to recognize that the performance and output of any enterprise depends entirely upon the quality of the human organization and its capacity to function as a tightly knit, highly motivated, technically competent entity. High educational efforts are not accomplished by impersonal equipment and computers. These goals are achieved by human beings. Successful organizations are those making the best use of individuals to perform well and efficiently all the tasks required to accomplish the aims and objectives for which organizations exist.[26]

The theme of this paper imposes the goal of changing the organizational accomplishments -- as related to educational accomplishments of LEP students. Halpin suggests that changes in the organization's accomplishments are the best criteria of the administrator's effectiveness.[27] Culbertson added that the capacity to cope constructively with change is the important test of leadership.[28] Referring to such change Lonsdale suggests that organizations need flexibility to accommodate to disturbances and to initiate new structures or to revise the goals of the organization.[29]

Values as they relate to organizational phenomena contribute to the quality of outcomes and changes. Blau described the integrative bonds of an organization as:

> the common values and norms...and the network of social relations in which processes of social interaction become organized.[30]

Teachers, by the nature of their jobs, become educational administrators. Teachers, administrators, students and others are all part of the social organization of the educational "system" operating in any community.

Communities and schools must practice the art of inclusion. The education and social needs of the LEP students must be met by the organized community that supports the work of the schools. The school administrators and teaching staff must meet the needs of LEP students.

It is needful to review a couple of management styles to reflect on possible strategies to include LEP related issues into every school organization in America.

Likert asserts that primarily two systems of management with different emphases developed side by side. The "job organization" system relies basically on the economic motives of buying a man's time and then telling him precisely what to do, how to do it, and at what level to produce. The "cooperative-motivation" system tends to use the principles and methods of scientific management and related management principles to a degree. This system taps not only the economic motives but additionally other strong motives, such as the ego motive.[31] He attempted to include the desirable features of each into an integrating principle of management which states that:

> The leadership and other processes of the organization must be such as to ensure a maximum probability that in all interactions and all relationships with the organization each member will, in light of his background, values, and expectations, view the experience as supportive and one which builds and maintains his sense of personal worth and importance.[32]

The basic principle of Likert's approach is that of "supportive relationships." He included four systems identified as: (1) exploitive authoritative; (2) benevolent authoritative; (3) consultative; and (4) participative.[33] He concluded that system four, "participative," is the most desirable, because as organizations move toward this system, the more productive and satisfying they become.

Several investigators, recognizing the relationship of values with human and interpersonal needs, have formulated classification schemes for these needs. Schutz's theory of interpersonal behavior proposes that each individual has three interpersonal needs: (1) inclusion, (2) control and (3) affection. His theory suggests:

> The term "interpersonal" refers to relations that occur between people as opposed to relations in which at least one participant is inanimate. It is assumed that, owing to the psychological presence of other people, interpersonal situations lead to a behavior in an individual that differs from the behavior of the individual when he is not in the presence of other persons.[34]

The interpersonal need of inclusion is behaviorally defined as the need to establish and maintain a satisfactory relation with people with respect to interaction and association. This is further defined as the need to establish and maintain a feeling of mutual interest with other people. This includes (1) being able to take an interest in other people to a satisfactory degree and (2) having other people interested in the self to a satisfactory degree. With regard to the self-concept, the need for inclusion is the need to feel that the self is significant and worthwhile.

The interpersonal need for control is behaviorally defined as the need to establish and maintain a satisfactory relation with people with respect to control and power. This is further defined as the need to establish and maintain a feeling of mutual respect for the competence and responsibility of others. This includes (1) being able to respect others to a satisfactory degree and (2) having others respect the self to a satisfactory degree. With regard to the self-concept, the need for control is the need to feel that one is a competent, responsible person.

The interpersonal need for affection is behaviorally defined as the need to establish and maintain a satisfactory relation with others with respect to love and affection. At the feeling level the need for affection is defined as the need to establish and maintain a feeling of mutual affection with others. This feeling includes (1) being able to love other people to a satisfactory degree and (2) having others love the self to a satisfactory degree. With regard to the self-concept, the need for affection is the need to feel that the self is lovable.

Schutz developed his efforts from the work of personality theorists. Of significance to his efforts was the work of Horney, Fromm, and Freud. Each of these identified three types or areas of interpersonal needs. Although the terminology is not identical in the descriptions of these areas, the definitions are quite similar. Horney identifies these areas as (1) moving toward people, (2) moving against people, and (3) moving from people.[35] Fromm identifies the areas as (1) withdrawal destructiveness, (2) symbiotic, and (3) love.[36] Freud identifies the three major systems as (1) erotic, (2) obsessional, and (3) narcissistic.[37]

Argyris suggests a four-dimensional classification including (1) inner needs and outer needs; (2) conscious and unconscious needs; (3) social needs; and (4) physiological needs.[38] Maslow developed his hierarchy of needs including five categories. In ascending order these are: (1) physiological needs; (2) safety needs; (3) belongingness and love needs; (4) esteem needs; and (5) the need for self-actualization. A basic part of this theory is that other and higher needs emerge when lower needs are satisfied, but not until they are satisfied.[39] The contribution of values both to individual and organizational behavior is commonly accepted by these organizational theorists. Parsons suggests that values are internalized cultural standards, norms, and expectations that influence a person's behavior. While value systems are highly personal, they are also involved in and affect the organization to which one holds membership. Parsons states this as: "A personal value system is in the social context, the network of rights and obligations in which an individual's value-commitment involves him in his social situation."[40] This would suggest that within the social systems context the individual's value orientations influence his perception of organizational components.

Value orientations develop through many ways. Education and training are important components of developing individual and community value orientations, or ethic constructs.

A paradigm shift is a change of how a person views reality. Education has the power to change personal views of reality. Education can increase perceptual acuity of teachers and administrators in working with all students, and particularly with LEP students.

School administrators and teachers should be able to use the total resources that the education industry has available for making sure that every teacher of LEP children is prepared to provide appropriate instruction to that student's needs. Appropriate instruction makes it possible for a LEP student to advance academically to the expectations of the school and the community at large while learning English. Appropriate instruction must rely on the totality of the resources available within the community, and on the total quality support from the administration of the schools. Resources are obtainable and are usefully articulated into standard school practices through attention to acceptable principles of management and teaching. Newer management attitudes are developing with the use of the concepts of total quality management (TQM), the quality sciences and the voluntary development of missions and standards. The education industry lacks such devices to measure progress of the education enterprise toward accomplishing its missions, goals and community expectations. These tenets of newer management constructs are included in the proposed steps designed to accomplish the paradigm shift for providing a quality education for all LEP students in the United States.

Towards the New Paradigm

How can we, then, as professional educators, accomplish a paradigm shift in teacher training for teachers of all LEP students in the United States? We will need to:

1. **Identify and keep what is good (what works) that we have learned, nationally, in working with LEP students, whether in large groups, small groups, or individually.**

There is no one best way to help LEP students achieve quality schooling. We need, as a profession, to continually contribute what we have learned as individuals and collectively and in working with LEP students. We need to use all the resources at our disposal in doing this and stretch ourselves to make sure a solid, accessible knowledge base is organized and immediately available for all education personnel, the community at large and parents, particularly.

2. **Develop acceptable levels of knowledge about what works by subject area as outlined in AMERICA 2000, but specifically for LEP students.**

AMERICA 2000 has set five core subject areas as those to be tracked for improvement in American education. They are mathematics, science, history, English, and geography. The improvement strategy for the nation will fail if the education of LEP students fails. Consequently, we professional educators, working on programs and practices for LEP students, must develop describable and specific programs for LEP students in each of the five core subject areas.

3. **Engage selected professorial and administrative persons from teacher training institutions in a national dialogue on numbers one and two above, through a series of coordinated symposia and workshops.**

Literature searches keep the profession alert to new developments, but usually much later than would be appropriate in a fast changing environment. We need to be sponsoring and holding a series of coordinated serious symposia, workshops, and developmental strategy sessions on each of the areas identified through the activities of one and two above.

4. **Provide general seminars for all college level education professors to learn administrative and teaching knowledge specifically appropriate for their content areas for working with LEP students.**

Periodically, especially during summers and other academic slow times, national seminars and conferences should be held to challenge the profession to develop the new paradigm and outline it, and use its information and knowledge base as it emerges.

5. **Provide ongoing help, nationally, for all professorial persons to build continually the knowledge that emerges from steps one through four above into teacher training curricula as appropriate at the local level.**

A national coordinated strategy, such as a national voluntary standards development activity, should be initiated so that all those who would be affected by step numbers one through four above might participate and gain from the knowledge base as it is being set into the new paradigm.

6. **Develop strategies for measuring the inclusion into teacher training appropriate curricula for the teaching of the knowledge about LEP learning needs and strategies.**

The application of the quality sciences to education provides guidance on developing and using appropriate measuring devices for ascertaining quality both from the perspective of the supplier of the services and the recipients of them. Quality can be measured and it is up to our profession to decide what to measure, how to do it, and who does it. Total Quality Management is one of the strategies of the quality sciences.

7. Learn the empowering process at the institutional level to provide the specific training for skills and knowledge to satisfy the needs of LEP students.

Institutional change within individual higher education institutions can occur either rapidly or slowly, depending on the environment of the moment. Leaders come and go, and bring with them their own perceived priorities and take away with them some of the momentum of special areas of interest that were alive and well as long as the leader was present. But, aside from the influence of individuals, each state has regulatory and governance issues that control and balance the operations and output of IHEs. It is extremely important to know how the regulatory and governance processes work at both state and institutional levels. State offices usually address general policies and local institutions concentrate on specific programs within general policy guidelines. Professional educators seem the most vulnerable to change in personnel through changes in operations policy, while professionals, i.e., professors, are generally viewed as experts who should be on target with issues in their field. Our specific challenge is to make sure the general policies of the state and the operational institutional policies are constructed to align with the critical issues in the professional fields. I suggest to you that the educational outcomes for LEP students, all LEP students, is a critical issue in American education.

What is before us is a significant challenge -- but a challenge that is attainable. Americans have a history of meeting challenges. We can meet this one also, if we successfully collaborate in such ways so as to benefit LEP students from our cooperative synergy. We must have total quality cooperation of all education professionals who are aware of the issues involved and are totally committed to their solution. We can do it and we can do it more quickly and easily if we involve all those who would be affected by our actions at the start. Let's move it on TOGETHER so the LEP train can reach its destination.

Notes

[1] National Commission on Children, Beyond Rhetoric A New American Agenda for Children and Families: A Final Report of the National Commission on Children (Washington, DC: Government Printing Office, 1991).

[2] International Association for the Evaluation of Educational Achievement, The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective (Champaign, IL: Stipes, 1987), pp.14-29.

[3] International Association for the Evaluation of Educational Achievement, Science Achievement in Seventeen Countries: A Preliminary Report (Oxford: Pergamon Press, 1988), pp. 25-43.

[4] U.S. Department of Education, Office of Elementary and Secondary Education, Preparing Young Children for Success: Guideposts for Achieving Our First National Goal, 1991, pp. 4-5.

[5] G. Natriello, E.L. McDill, and A.M. Pallas, Schooling Disadvantaged Children: Racing Against Catastrophe (New York: Teachers College Press, 1990), pp. 25-28.

[6] Ibid.

[7] Robert Milk, "Preparing Teachers for Effective Bilingual Instruction," in Languages in School and Society: Policy and Pedagogy, Mary E. McGroarty and Christian J. Faltis, eds. (Berlin and New York: Mouton de Gruyter, 1990), pp. 267-280.

[8] Milk, "Preparing Teachers for Effective Bilingual Instruction," p. 270.

[9] Milk, "Preparing Teachers for Effective Bilingual Instruction," p. 270.

[10] Milk, "Preparing Teachers for Effective Bilingual Instruction," pp. 273-274.

[11] Eleanor Thonis, "Competencies for Teachers of Language Minority Students," in Languages in School and Society: Policy and Pedagogy, Mary E. McGroarty and Christian J. Faltis, eds. (Berlin and New York: Mouton de Gruyter, 1990), pp. 282-283.

[12] U.S. Department of Education, Office of Elementary and Secondary Education, Preparing Young Children for Success, p. 2.

[13] U.S. Department of Education, Office of Elementary and Secondary Education, Preparing Young Children for Success, p. 3.

[14] U.S. Department of Education, Office of Elementary and Secondary Education, Preparing Young Children for Success, p. 4.

[15] S. Purkey and M.S. Smith, "Effective Schools: A Review," Elementary School Journal 83(1983):427-452; J.E. Chubb and T.M. Moe, Politics, Markets, and America's Schools (Washington, DC: Brookings Institution, 1990), pp. 159-166.

[16] Purkey and Smith, "Effective Schools," pp. 443-446; M.S. Smith and J. O'Day, Systemic School Reform: Draft (Palo Alto, CA: Stanford University School of Education, 1990), pp. 3-5.

[17] Purkey and Smith, "Effective Schools," pp. 440-442; Smith and O'Day, Systemic School Reform, p. 4.

[18] Smith and O'Day, Systemic School Reform, p. 4; J.S. Coleman and T. Hoffer, Public and Private High Schools: The Impact of Communities (New York: Basic Books, 1987), pp. 3-27.

[19] Rosemarie V. Rosen, ". . . For the Short and Long Term . . ." Harvard Educational Review 54 (February 1984):27.

[20] Bonnie Benard, Barbara Fafoglia, and Jan Perone, "Knowing What to Do--and Not to Do--Reinvigorates Drug Education," ASCD Curriculum Update (February 1987): 2.

[21] For example, as featured in the "Excellence: School by School" issue of Educational Leadership (March 1985) and the "Partnership: Building Links Between Schools and Communities" of Phi Delta Kappan (February 1984).

[22] Beatrice A. Ward and Joseph T. Pascarelli, "Networking for Educational Improvement," in The Ecology of School Renewal, John I. Goodlad, ed. (Chicago: University of Chicago Press, 1987), p. 194.

[23] Richard L. Andrews, "The School-Community Interface: Strategies of Community Involvement," in The Ecology of School Renewal, John I. Goodlad, ed. (Chicago: The University of Chicago Press, 1987), p. 158.

[24] Charles D. Pauly, "Meeting Goals is Easier When Everyone Pulls in the Same Direction," The American School Board Journal (November 1987):52.

[25] R.C. Lonsdale, "Maintaining the Organization in Dynamic Equilibrium," in Behavioral Science and Educational Administration, Sixty-third Yearbook of the National Society for the Study of Education, Part II, D. Griffiths, ed. (Chicago: University of Chicago Press, 1964), pp. 142-177.

[26] R. Likert, The Human Organization: Its Management and Value (New York: McGraw-Hill Book Company, 1967), p. 134.

[27] A.W. Halpin, Theory and Research in Administration (New York: The MacMillan Company, 1966), p. 50.

[28] J.A. Culbertson, "The Preparation of Administrators," in Behavioral Science and Educational Administration, p. 315.

[29] R.C. Lonsdale, p. 174.

[30] P.M. Blau, "Structural Effects," American Sociological Review, Vol. 25, No. 2 (April, 1960), p. 178.

[31] R. Likert, New Patterns of Management (New York: McGraw-Hill Book Company, 1964).

[32] R. Likert, The Human Organization, p. 47.

[33] R. Likert, New Patterns of Management, pp. 222-234.

[34] W.C. Schutz, FIRO: A Three-Dimensional Theory of Interpersonal Behavior (New York: Holt, Rinehart and Winston, Inc., 1960), p. 14.

[35] K. Horney, Our Inner Conflicts (New York: W.W. Norton and Company, 1945), pp. 40-43.

[36] E. Fromm, Man for Himself (New York: Holt, Rinehart and Winston, Inc., 1947), pp. 109-110.

[37] S. Freud, "Libidinal Types," in Collection Papers, Vol. 5 (London: Hogarth, 1950), pp. 247-248.

[38] C. Argyris, Personality and Organization (New York: Harper and Brothers, 1957), pp. 33-41.

[39] A.H. Maslow, Motivation and Personality (New York: Harper and Row Publishers, 1954), pp. 80-92.

[40] T. Parsons, Structure and Process in Political Systems (Glencoe: The Free Press, 1960), p. 175.

References

Alatis, J., Stern, H., & Strevens, P. (Eds.). (1983). Applied linguistics and the preparation of second language teachers: Toward a rationale. Georgetown University Round Table on Languages and Linguistics. Washington, DC: Georgetown University Press.

Argyris, C. (1960). Personality and organizations. New York: Harper Brothers.

Blau, P. M. (1960). Structural effects. American Sociological Review, 25(2).

Celce-Murcia, M. (1983). Problem solving: A bridge builder between theory and practice. In J. Alatis, H. Stern, & P. Strevens (Eds.), Georgetown University Round Table on Languages and Linguistics (pp. 97-105). Washington, DC: Georgetown University Press.

Clark, E., & Mike, R. (1984). Training bilingual teachers: A look at the Title VII graduate in the field. NABE Journal, 8(1), 41-53.

BEST COPY AVAILABLE

Collier, V. (1985). University models for ESL and bilingual teacher training. Issues in English language development (pp. 81-90). Rosslyn, VA: National Clearinghouse for Bilingual Education.

Culbertson, J. A. The preparation of administrators. Behavioral Science and Educational Administration.

Freud, S. (1950). Libidinal types. In Collected Papers Vol. 5. London: Hagarth.

Fromm, E. (1947). Man for himself. New York: Holt, Rinehart and Winston.

Guba, E. G. & Bidwell, C. E. Administrative relationships. (Monograph No. 4). Chicago: University of Chicago, Midwest Administration Center.

Halpin, A. W. (1966). Theory and research in administration. New York: MacMillan.

Horney, K. (1945). Our inner conflicts. New York: W. W. Norton & Co.

Likert, R. (1964). New patterns of management. New York: McGraw Hill.

Likert, R. (1967). The human organization: Its management and value. New York: McGraw Hill.

Lonsdale, R. C. (1964). Maintaining the organization in dynamic equilibrium. In Behavioral science and educational administration. Sixty-third yearbook of the National Society for the Study of Education, Part III. Chicago: University of Chicago Press.

Maslow, A. H. (1954). Motivation and personality. New York: Harper & Row.

McGroarty, M., & Galvin, J. (1985). Culture as an issue in second language teaching. In M. Celce-Murcia (Ed.), Beyond basics: Issues and research in TESOL (pp. 81-95). Rowley, MA: Newbury House.

McKeon, D. (1985). Some common components in training bilingual, ESL, foreign language, and mainstream teachers. Issues in English Language Development (97-106). Rosslyn, VA: National Clearinghouse for Bilingual Education.

Mellgren, M., Walker, C., & Lange, D. (1988). The preparation of second language teachers through post-baccalaureate education. Foreign Language Annals, 21, 121-129.

Milk, R. (1985). The changing role of ESL in bilingual education. TESOL Quarterly, 19(4), 657-672.

Milk, R. D. (1991). Preparing teachers in effective bilingual instruction. In M. E. McGroarty & C. J. Faltis (Eds.), Languages in school and society: Policy and pedagogy. New York: Mouton de Gruyter.

Parsons, T. (1960). Structure and process in political systems. Glencoe: The Free Press.

Politizer, R. (1978). Some reflections on the role of linguistics in the preparation of bilingual/cross-cultural teachers. Bilingual Education Paper Series, 1(12). (ERIC Document Reproduction Service No. ED 161 274)

Schutz, W. C. (1958). FIRO: A three-dimensional theory of personal behavior. New York: Holt, Reinhart, & Winston.

Soto, L. D. Understanding bilingual/bicultural young children. Young Children, 42(2), 30-36.

Thonis, E. W. (1991). Competencies for teachers of language minority students. In M. E. McGroarty & C. J. Faltis (Eds.), Language in school and society: Policy and pedagogy. New York: Mouton de Gruyter.

Response to John Steffens' Presentation

Virginia Collier
George Mason University, Virginia

I heartily affirm the challenge that John has presented to our field, to reach all teachers and administrators in the United States and to provide them with the appropriate training to work with the few or the many limited English proficient students that they may receive in their classrooms and schools. I would like to extend the idea to include not just limited-English-proficient students, but all language minority students. SE should be more consistent with the original Title VII Bilingual Education Act, which addressed not only students of limited English proficiency but all language minority students, knowing that all language minority students, even those who are fluent only in English, still need help.

As we language minority educators approach the challenge, I believe that the key to the most practical solution to developing a strategy for reaching all teachers and administrators is to link up with the current school reform movement taking place across the country. There are a lot of exciting things happening. Some of the major reforms taking place have to do with the administrative structure of schools, such as changes to make the decision-making process more collaborative for all participants -- including teachers, students, parents, administrators, and community. Think about what that means for language minorities; it means parent involvement in a way not possible before. Another major change in the administrative structure of schools is the movement toward eliminating tracking, which was a side effect of our efforts at compensatory education reforms of the 60s and 70s and has had disastrous effects on all minority students. Jeannie Oakes and others have spoken eloquently on this issue.

Other major changes currently taking place in schools are focused on the curriculum and methods of teaching such as: first, the development of higher order thinking skills, including hands-on experiential learning and problem solving; second, team teaching; third, the more meaningful integration of all subject areas as a result of the teaming; fourth, whole language approaches to teaching language, including teaching writing as a process and getting students to write a great deal; and fifth, the use of cooperative learning and the consequent elimination of ability grouping, another form of tracking.

Our research on language minority education to date indicates that all of these promising practices also help language minority students significantly. A monograph written by Lorraine Valdez-Pierce, Effective Schools for Language Minority Students, published by the Mid-Atlantic Equity Center in 1991, examining the school reform movement and effective strategies for language minority education that are connected to the changes now taking place in schools.

I believe that our field has made some mistakes in the past by falling into the same trap that special education got into in its early stages of development through the creation of separate programs and classrooms for the special needs of students. Over the last decade, in the 80s, special education has worked very hard to mainstream students who were formerly placed in special education classes, to find the least restrictive environments for students with special needs. This has involved some creative team teaching, getting special education teachers and mainstream teachers back together again, moving students back into integrated programs designed for all students.

We language minority educators must face the same issues. It is clear that schools by the end of this decade must eliminate tracking and ability grouping. This means a total restructuring of the secondary school. We have a long way to go on this issue. Middle schools are doing this right now; teaming is in; meaningful integration of subject matter is taking place. We language minority educators must join these reform efforts now and make sure that the decisions for new school structures reflect the needs of language minority students. The amazing thing is that many of these reform efforts do reflect best practice for the education of language minority students, but what is missing from the teacher and administrator training currently going on is a clear synthesis of the research on bilingualism and biculturalism and how a student's two languages and cultures interact with and influence the process of learning. We have to work on finding a way for mainstream administrators and teachers to get this information.

Our first step to creating a new paradigm that John recommends for teacher training might be to gather together the most meaningful syntheses of research on language minority education and make them readily available to all teacher trainers. The Center for Cultural Diversity and Second Language learning has been given the responsibility for publishing some of these syntheses. The rest of us can also be working on dissemination of research syntheses in our publications.

One possible means for dissemination of this knowledge base would be the institutes for IHE education faculty that John has proposed in his paper. This could become a trainer model and it seems very exciting. Those attending the institutes would be given procedures and ideas for retraining their own faculty when they return to their institutions. However, special education has already tried some of these kinds of institutes with somewhat limited success. In my own experience with my colleagues in higher education, I find that it is important to find some kind of very clever institutional incentives for faculty retraining; otherwise, they will go their own independent ways and do things as they have always done them.

I would like to share with you a model that we are exploring at George Mason University. Up to this point in our teacher training program, special education faculty have trained special education teachers; bilingual education faculty have trained bilingual and ESL teachers, and mainstream faculty have trained mainstream teachers, for the most part. There is some course work which all three groups of preparing teachers attend jointly, but there are many special courses for the specialists. Yet, special education and bilingual education faculty have felt increasingly separated from mainstream faculty. While we share decisions across all education faculty, we have fallen into the same segregated institutionalization of our fields that has occurred in public schools.

We have decided that we must change this pattern. Since the school reform movement is pushing for lots of team teaching at elementary school and middle school levels, and I hope someday this will also be a teaching pattern in secondary schools, we faculty feel that we should model teaming by faculty teaming in our teacher training program. We are just beginning to explore the idea. This will involve lots more preparation time, with both faculty members attending, but all class sessions will allow faculty to learn from each other and to incorporate language minority and special education issues into all teacher training courses, in an integrated program.

As we are talking, we find that we agree on the major knowledge that we want to get across to teachers and each of us has special expertise to contribute to the courses that the other faculty respect as important for preparing teachers to know. We expect this teaming to enrich our own knowledge and skills. We are thinking that the curricular and instructional reform now taking place in many schools will become the cornerstone of our teacher training program: teaching higher order thinking skills, experiential/interactive learning, whole language approaches, integration of language and content across the curriculum, use of cooperative learning and elimination of tracking and ability grouping, and, added to that, understanding bilingualism and multiculturalism and all of the dynamic aspects of linguistic and cultural process taking place inside and outside the classroom. A quote from Lorraine Valdez-Pierce's book provides an example of training strategies needed in our teaming: "Recent research suggests that transmission models of education are not effective with minority students who are at-risk of failure in schools....For these students, reciprocal interaction models based on student collaboration have been shown to be more effective....These define the teacher's role as that of a facilitator, one who makes things happen by providing a learning environment which promotes student interaction and efficient questioning strategies necessary to the development of higher order skills" (Valdez-Pierce, 1991, p. 20).

Perhaps this is just one way of initiating development of a new paradigm. I know there are others. In some states the pressure of student demographic changes, with increasing language minority needs, will force teacher training faculty to seek change. A major change agent can be changes in certification standards for mainstream teachers, which Rosita will address next, explaining changes taking place in California.

I would like to finish my comments by addressing the issue of the training of bilingual and ESL teachers more specifically. As we watch and join these reform movements for all education, we must speak out to clarify that language minority education should not follow the outdated notions of compensatory and remedial education. Basic skills approaches are a sure way to keep our students at the bottom of the success ladder. We must demand high quality training for bilingual and ESL teachers, integrated with mainstream teachers, that keeps up with the latest research on what works with all students. Bilingual students want to be active learners; they want to have access to all the advantages provided for gifted and talented learners.

As we look at ways to integrate all learners into meaningful classes, we must continue to expand ways for providing support for language minority students' cognitive development in their first language. Research clearly shows that first language cognitive development is crucial to second language academic achievement. There are many meaningful ways to support the first language, through the school environment and attitudes toward the first language, through family education in the school evenings and weekends, through encouragement of parents' first language activities with children at home, and (the best of all possible worlds from my point of view) through two-way bilingual programs where English speakers respect and share in the process of learning a second language.

We cannot implement two-way bilingual schools everywhere, but even in neighborhoods where there are just a few limited-English-proficient students, when English speaking parents want their children to learn the first language of those limited-English proficient students, a two-way program can be perceived by all as a gifted and talented class with the highest expectations for success. I'm currently watching the changes that are taking place in parent attitudes occurring in Fairfax County Public Schools here in our metropolitan area, where the eight bilingual schools now in their third year of implementation have incredible parent support, with many other parents clamoring for similar programs in their schools. There are only a few language minority students in these classes because there are just a few in each of these schools. Those language minority students are benefitting enormously from the prestige suddenly given to their language and the pride and self-esteem they feel. They are doing very well academically along with their English-speaking peers.

One more example is my daughter's own two-way bilingual school in the District of Columbia Public Schools. I conducted a small case study a couple of years ago, contacting all the Hispanic and Anglo graduates that I could locate from the first year of implementation of the program in 1971. All 20 that I found are now college graduates who have continued full use of their two languages in their careers. They are very successful professionals, and the most amazing thing is that many of the Anglo as well as Hispanic students have chosen social service professions including teaching (some of them are bilingual teachers), and they are assisting language minority communities with successful achievement and upward mobility. I hope we can keep this in mind as an ideal vision of integrated, exciting schooling for the future of all our students.

Response to John Steffens' Presentation

Rosita G. Galang
University of San Franciso

In the last two years, in light of the demographic changes in the California school-age population, the Bilingual Cross Cultural Advisory Panel of the Commission on Teacher Credentialing (CTC) has been engaged in the re-examination of the existing preparation programs, credentials, certificates, and examinations for teachers of students from diverse linguistic and cultural backgrounds. As a member of this panel and as a faculty in an Institute of Higher Education (IHE) involved in the preparation of teachers in a state where there were more than 860,000 identified "limited English proficient" (LEP) students representing 137 languages in the 1989-90 school year, I am deeply interested in the topic of today's session. Therefore, I am grateful to OBEMLA for giving me the opportunity to learn from the session this afternoon and also to share my thoughts and those of my colleagues on the panel regarding the preparation of teachers of LEP students.

As some presenters in an earlier session and the participants in last year's symposium have pointed out, the term LEP is not acceptable to many who regard it as demeaning, derogatory, and/or focusing on students' limitations rather than potential. Although I would much rather use a different term such as beginning English learners or potentially English proficient students, I will use the term LEP since it's the term used in this symposium and the paper to which I have been invited to respond.

As a discussant with only twenty minutes to respond to the paper, I will limit my comments to these areas: the need for a paradigm shift, a suggested paradigm for the preparation of teachers, and steps that could be taken to accomplish the said paradigm shift.

Specifically, my response aims to do the following:

1. Point out selected assumptions and concepts presented by the author that I generally agree with and therefore form the bases of my comments.

2. Present some of my reflections regarding the preparation program described in the paper.

3. Suggest a paradigm with the potential of meeting the need for trained teachers of LEP students.

4. Give some reflections on the steps that might be taken to accomplish the paradigm shift.

Need for a Paradigm Shift

From the assumptions and concepts presented in the paper, I have selected a few as bases of my brief response. These are:

1. That language minority children, many of whom are identified as limited English proficient, make up a growing proportion of our student population and their rapid increase in number is expected to continue. In fact, in California, their rate of increase and extent of diversity have grown in recent years.

2. That current LEP students are not concentrated in any one location but, instead, reside in three types of communities and consequently study in three types of classrooms.

   Type A -- where there are significant numbers of LEP students of similar ethnolinguistic backgrounds

   Type B -- where there are significant numbers of LEP students of different ethnolinguistic backgrounds

   Type C -- where there are small groups of LEP students of different ethnolinguistic backgrounds and are sparsely distributed so that only one or a few might be found in the classrooms

3. That it is our responsibility to provide equal educational opportunities for all LEP students, even if there's only one or two in the classroom.

4. That we haven't been able to supply enough bilingual teachers to teach in classrooms where there are concentrations of LEP students of similar or different ethnolinguistic backgrounds (Type A and B classrooms).

   Corollary to this assumption is the need to continue the training of bilingual teachers. The number of teachers who have the necessary instructional, linguistic, and cultural competencies have not kept pace with the continued growth and diversity of the language minority student population.

5. That we have not paid attention to and therefore need to look at the LEP students sparsely distributed in classrooms (Type C).

6. That we need a paradigm shift in the preparation of teachers of LEP students.

Building a New Paradigm

While the change in the preparation program described or implied in the paper is a commendable attempt to respond to the demand for teachers of LEP students, it can only be considered as a short-term solution to the shortage of the needed teachers. I should point out that its focus on what all teachers need to know and be able to do when a few LEP students are assigned to their classrooms is a step in the right direction. However, its lack of connection or statement of connection to the preparation of English as a Second Language (ESL) and bilingual teachers makes it an inadequate change. At best, the products of such a program are prepared to teach in Type C classrooms--those with LEP students who are sparsely distributed. It cannot account for the preparation of teachers needed in Types A and B classrooms.

Historically, the preparation of ESL, bilingual, and the so-called "regular" teachers have been designed and implemented separately and independently of each other in response to specific needs at particular times. Perhaps our inability to meet the demand for teachers that could function in the three types of classrooms can partly be attributed to this unfortunate situation. The author points out that we need a new paradigm in the preparation of teachers of LEP students. I agree, and I strongly believe that we need a paradigm that relates the preparation of teachers in a comprehensive system for LEP students.

Prerequisite to the conceptualization of such a paradigm is the examination of the instructional needs of LEP students whether they are in Classroom A, B, or C. LEP students, like all students, need opportunities to learn the core curriculum. Traditional or mainstream instruction in English denies them access to the core curriculum. Therefore, their basic instructional needs are English language development and access to the core curriculum. English language development involves ESL instructional methodologies and access to the curriculum involves academic instruction in the primary language and specially designed academic instruction in English. Specially designed academic instruction in English may be defined as the teaching of the content of the core curriculum in English to LEP students in a way that considers their level of English proficiency, for example, through sheltered English subject matter instruction. Here the teacher utilizes instructional modifications such as simplified speech, and the use of verbal clues to make the language comprehensible to the students. This type of instruction is used where primary language instruction is not possible or available.

The instructional needs of LEP students can be met by using a bilingual teacher or a team of teachers who can provide ESL instruction and bilingual instruction (Primary Language and English instruction). Unfortunately, these are not always feasible, practical or advisable for several reasons. The continuing shortage of bilingual teachers and the increasing linguistic and cultural diversity of our student population emphasize the need for teachers who are prepared to provide English language development instruction and equal access to the core curriculum. Where primary language instruction is not a viable option, specially designed academic instruction in English is accepted as an alternative.

> In Type A classrooms, ESL instruction and primary language instruction can be provided by bilingual teachers.
>
> In Type B classrooms, ESL instruction and specially designed academic instruction in English can be provided by a Language Development Specialist.
>
> In Type C classrooms, instruction in English can be provided by "regular" teachers who have been trained in multicultural education.

The paper focuses on what all teachers need to know and be able to do when LEP students are assigned to their classrooms, specifically in small numbers. The preliminary list of content areas cited in the paper could serve as core training areas for all teachers, bilingual or non-bilingual, and may be considered as the first level or component of the new paradigm. Teachers prepared in these content areas, usually called "regular" teachers, may serve in Type C classrooms. The second level may include the said core training plus training in ESL instruction and specially designed academic instruction in English. Teachers prepared by such a program, identified as Language Development Specialists (LDS) (for lack of a better term), may be assigned to Type B classrooms. The third level may include the same core training, training in ESL and specially designed academic instruction in English, and the following: development of proficiency in the student's primary language, increased knowledge of the student's background culture, and skills in teaching the primary language and using it as a medium of instruction. Teachers prepared by this program, known as bilingual teachers, may be assigned to Type A classrooms. It should be pointed out that depending on the needs of the students, the bilingual teachers are also prepared to teach in all types of classrooms while the LDS are also prepared to teach in Type C classrooms.

In California, the Commission on Teaching Credentialing Standards for Teacher Preparation Programs have already been revised to include multicultural education and second language acquisition as part of the preparation of all Multiple Subjects (Elementary) and Single Subject (Secondary) teachers. Still, the standards are being reexamined to further strengthen or increase the emphasis in the two areas. These are the "regular" teachers who serve in Type C classrooms.

Just last August, the conceptual framework proposed by the CTC Bilingual Cross-Cultural Advisory Panel was accepted by the Commission. The said framework exemplifies the paradigm that I have just described.

Matching the types of instruction needed in classrooms with significant numbers of LEP students (Type A and B classrooms), three types of credential/preparation programs/examinations are included in the framework:

1. Multiple Subjects/Single Subject Credential with a Cross-Cultural Language and Academic Development Emphasis (CLAD) which, in addition to the core training for "regular teachers" includes training in these areas:

    a. Language Structure, Acquisition and Development
    b. Bilingual and ESL Models and Methodology
    c. "Generic" Culture or Cross-Cultural Communication

2. Multiple Subjects/Single Subject Credential with a Bilingual Cross-Cultural Language and Academic Development Emphasis (BCLAD) includes the training for the "regular teachers," the training for the CLAD teachers, and preparation in three additional areas:

    a. Methodology for Instruction in the Language of Emphasis
    b. The Culture of Emphasis
    c. The Language of Emphasis

3. Culture and Language Specialist Credential includes preparation for the CLAD Credential holder plus further preparation on

    a. Assessment
    b. Curriculum Development
    c. Staff Training
    d. Community/Parent Relations

4. Bilingual Culture and Language Specialist Credential includes preparation for the BCLAD credential holder plus further preparation on the same areas cited in 3.

The California theoretical framework relates the preparation for teachers of LEP students to the instructional needs of LEP students in the three types of classrooms described earlier.

All teachers (including the "regular teachers") will be able to provide instruction in Type C classrooms since everyone will have the "core training" needed to be prepared to deal with one or a few LEP students.

CLAD teachers will also be able to provide instruction in Type B classrooms where significant numbers of LEP students of different ethnolinguistic backgrounds will receive instruction in ESL and specially designed academic instruction in English.

BCLAD teachers will be able to provide instruction in Type A classrooms where there are significant members of LEP students of similar ethnolinguistic backgrounds and therefore will receive instruction in and through the primary language, specially designed academic instruction in English, and instruction in ESL.

The Culture and Language Specialists will provide the leadership and resources needed by CLAD and BCLAD teachers.

The paradigm which I have described appears to be relevant and has the potential of being used as a guide in designing preparation programs for teachers of LEP students.

1) It provides a framework for the training of teachers who can serve in the three types of classrooms.

2) It shows the common areas shared by the preparation of the different teachers and the additional areas of training for the same.

3) It presents teachers with options for obtaining training in teaching LEP students depending on their goals and qualifications. For example, the monolingual English or "regular" teacher might start with the preparation for Type B classrooms and ultimately strive for the preparation for Type A classrooms.

4) It provides opportunities for integrating areas of bilingual education and ESL and content area instruction and therefore encourages collaboration among bilingual and non-bilingual teachers and their trainers.

Implementing the Paradigm Shift

In the last section of the paper, steps that need to be taken to accomplish the paradigm shift are listed and discussed briefly. Allow me to give my reflections on two of them.

Engage selected professorial and administrative persons from teacher training institutions in a national dialogue through a series of coordinated symposia and workshops.

While a national dialogue among professorial and administrative persons from teacher training institutions is needed, collaboration needs to be expanded. The value of collaboration in teacher education in general and language minority teacher education in particular cannot be overemphasized. Collaboration is critical at different levels and among everyone involved in and affected by the process--teachers, students, administrators, teacher trainers, and others. As Emily DiMartino wrote in Education in 1991, collaboration is a vertical phenomenon as elementary school children and teachers interact with personnel at the university level and horizontal as liberal arts and education faculty within the college work together to strengthen the training of prospective teachers. Collaboration should be an ongoing process during the planning, designing, implementing, evaluating, and reviewing or modifying steps.

Learn the empowering process at the institutional level to provide the specific training for skills and knowledge to satisfy the needs of LEP students.

I suggest that we also look at the empowering process in a light different from that discussed in the paper. In an article that appeared in the Harvard Educational Review in 1986, Alma Ada underscored that for teachers to be able to provide creative education for language minority students, they themselves need to experience the liberating forces of this type of education.

Teachers have to be empowered through an understanding of the societal forces that have influenced their linguistic and cultural identity so that they cease being passive and, instead, become pro-active in transforming their own selves and assuming a leadership role in the world around them. Through empowerment of teachers, we may expect empowerment of students. If successful programs for language minority students are those that empower students, that is, develop in them a strong sense of confidence in who they are and their ability to learn, then the empowering process should be an important component of the paradigm that will be used as a guide for designing preparation programs for teachers of LEP students.

Conclusion

The paradigm that I have just described was presented in response to the challenge posed in the paper regarding the need for a paradigm shift. The paradigm is by no means final and therefore may be modified as societal changes that affect education occur. Furthermore, it is not meant to dictate what teacher training should be but, instead, to guide the design of teacher preparation programs.

As we collaboratively build a paradigm that is responsive to the demand for teachers of LEP students, let's keep in mind that our ultimate goal is to prepare teachers who can provide equal and quality educational opportunities for linguistically and culturally different students.

Educational Research and Teacher Training for Successfully Teaching Limited English Proficient Students

Carl A. Grant
University of Wisconsin, Madison

The education of our nation's teachers is of vital importance to all of us. Unfortunately, the research data essential to making informed decisions about how to structure teacher education is meager. As the editors of the Handbook of Research on Teacher Education (1990) observed in their preface:

> ...there has been notable recent progress, but the research basis for such important work as educating the nation's teachers is still extremely thin. Although the importance of research is being espoused, little progress is being made. (p. ix) (my emphasis)

One reason for the "extremely thin" research in teacher education is that few university teacher educators do research beyond their doctoral dissertation. Another is that many of the research findings are of the "I believe" type, based on questionnaires or surveys, with conclusions that exceed the limits and power of the instruments used. The variables studied tend to be those that are convenient, rather than those that have the potential of making a difference.

Much of the teacher education literature is filled with descriptions of programs rather than careful analysis of program outcomes. Most studies are singular, one time occurrences that are difficult to connect to any other studies in the field. The lack of understanding about school and classroom life often leads to inappropriate or ill conceived research paradigms. Finally, most educational institutions usually are not committed to using research as a basis for policy making in teacher education. Instead, they are the most often committed to "doing it our way," with very little attention paid to research reports or the efforts of other institutions (Houston, Haberman & Sikula, 1990).

While the lack of a substantial body of solid research is a serious problem in teacher education in general, it is a doubly serious problem when it comes to research on the preparation of teachers to work in culturally diverse schools, especially when that preparation includes working with limited English proficient (LEP) students.

Grant and Secada (1990), in their analysis of research studies on preparing teachers for diversity, were only able to locate 23 appropriate research studies. These studies consisted of 16 that addressed preservice education and 7 that addressed in-service education. Seventeen of these studies were concerned with multicultural education; seven with gender equity; and **one** with second language issues. Three studies overlapped on multicultural education and gender issues.

Focus and Organization of this Paper

The purpose of this paper is to examine the research on teacher training particularly as it relates to preservice and in-service teacher preparation, of teachers to work with LEP students. It will highlight successful programmatic patterns and innovations based on research for preparing teachers to work with LEP students. A discussion of the criteria used to determine programmatic success will be presented.

Two analytic paradigms will be used to examine and evaluate teacher preparation programs. The first level of analysis of LEP teacher preparation programs will include the "Framework for Intervention for Empowering Minority Students" proposed by Cummins. Cummins (1988) argues that, "...a major reason previous attempts at educational reform have been unsuccessful is that the relationships between teacher and students and between schools and communities have remained essentially unchanged" (p. 18). His theoretical framework includes four areas that teacher training programs for LEP students need to address: (1) cultural/linguistic incorporation, (2) community participation, (3) pedagogy, and (4) assessment.

The second level of analysis of LEP teacher preparation programs will include the multicultural framework first proposed by Grant and Sleeter (1985). This framework will help in the interpretation of the kinds and quality of attention to language and cultural diversity in each program. The multicultural framework includes five approaches for dealing with race, class, gender and disability diversity in schools: (1) Teaching the Exceptional and Culturally Different, (2) Human Relations, (3) Single Group Studies, (4) Multicultural Education, and (5) Education That Is Multicultural and Social Reconstructionist.

The chapter is organized to include both preservice and in-service education together because of the paucity of research exclusively dealing with preservice teacher preparation for working with LEP student. The literature reviewed will be organized and discussed according to Cummins' (1986) theoretical framework. The literature reviewed will then be examined in terms of the approaches to diversity proposed by Grant and Sleeter (1985).

Next, a general discussion of the successful practices common to both preservice and in-service teacher education programs will be presented. Finally, a discussion that compares the research findings to the observations on research in teacher education offered by Houston, Haberman and Sikula (1990) will be presented. Before beginning, a discussion of the analytic paradigms is in order.

Two Analytical Paradigms

Cummins' Theoretical Framework for Examining LEP Teacher Education Programs

The central tenet of Cummin's (1986) framework "... is that students from 'dominated' societal groups are 'empowered' or 'disabled' as a direct result of their interactions with educators in the school" (p. 21). Cummins states, "These interactions are mediated by the implicit or explicit role definitions that educators assume in relation to four institutional characteristics" (p. 21). Cummins defines these four institutional characteristics as:

1. minority students' **language** and **culture** are incorporated into the school program;
2. minority **community participation** is encouraged as an integral component of children's education;
3. the **pedagogy** promotes intrinsic motivation on the part of students to use language actively in order to generate their own knowledge; and
4. professionals involved in **assessment** become advocates for minority students rather than legitimizing the location of the 'problem' in the student. (p. 21) (my emphasis)

A modification of the framework was made for this study. This modification uses these key concepts (language and culture, community participation, pedagogy and assessment) as they are more broadly defined and used in the educational literature. LEP teacher education programs are then examined to see if these key concepts are included in their program.

A Multicultural Topology For Classifying Studies

Grant and Sleeter (1985, 1989) And Sleeter and Grant (1987, 1988) argue that educators deal with race, class, language, gender, and disability diversity in schools in at least five different ways. Each of these ways or approaches provides an analysis of schools as institutions of society that have a history of discrimination on the basis of race, gender, class, and disability. Each approach offers a positive improvement over the Anglo-centric teaching that was for many years accepted as the status quo. However, each approach suggests its own way of improving schooling for the disfranchised.

The first of these approaches, Teaching the Exceptional and Culturally Different, helps fit people into the existing social structure and culture. Dominant traditional educational aims are taught by building bridges between the students and the school. The curriculum is made relevant to the students' background; instruction builds on students' learning styles and is adapted to their skill levels. Teaching culturally different or exceptional children accommodates such students by altering regular teaching strategies to match student learning styles through use of culturally relevant materials or remedial teaching strategies.

The Human Relations approach attempts to foster positive affective relationships among individuals of diverse racial and cultural groups, and/or between males and females, to strengthen students' self-concept and to increase school and social harmony. The human relations curriculum includes lessons about stereotyping and individual difference and similarities. Instruction includes the use of cooperative learning. Teacher education from a human relations perspective prepares teachers to honor diverse student backgrounds and to promote harmony among students. Unfortunately, real conflicts between groups are often glossed over in the effort.

The Single-Group Studies Approach promotes structural social equality for, and immediate recognition of, the identified group. Commonly implemented in the form of ethnic studies or women's studies, this approach assumes that knowledge about particular oppressed groups should be taught separately from conventional classroom knowledge, in either separate units or separate courses. Single-group studies seek to raise people's consciousness about an identified group, by teaching its members and others about the history, culture, and contributions of that group, as well as how the group has worked with the dominant groups in our society or has been oppressed by them.

The Multicultural Education approach promotes social equality and cultural pluralism. The curriculum is organized around the contributions and perspectives of different cultural groups, and pays close attention to gender and disability equity. Multicultural education builds on students' learning styles, adapts to their skill level, and involves students actively in thinking and analyzing life situations. This approach also encourages schools to include diverse racial, gender, and disability groups in their staffing patterns.

The Education That Is Multicultural and Social Reconstructionist approach extends the previous approaches by teaching students to analyze inequality and oppression in society, and by helping them to develop skills for social action. Education That Is Multicultural and Social Reconstructionist promotes social structural equality and cultural pluralism and prepares citizens to work actively toward structural equality. Having examined these analytic paradigms, let us begin the review of the literature.

Teacher Education Programs for Language Minority Students

Language and Culture

Cazden and Mehan, (1989) Diaz, (1987) and Mehan & Trujillo, (1989) discuss the need for teachers to understand the importance that language and culture have on student success. For example, Cadzen and Mehan (1989) argue that outcomes from the Kamehameha Early Education Program (KEEP) reported by project researchers (Au,1980; Voght, Jordan, & Tharp, 1987) and the work by Heath (1983) clearly indicate the significance of home culture and language to school learning. Cadzen and Mehan (1989) claim:

> A major question for teacher education is how to help teachers develop strategies to achieve such accommodations in a wide range of communities, including those with students from different cultures." (p. 54)

Mehan and Trujillo (1989) also point out that it is important that teacher educators know that "the connection between students' home and community knowledge and the demands of schooling are crucial for linguistic-minority students' school success" (p. 1). Mehan made the following comment during a discussion at the Linguistic Minority Research Project Conference held in 1988, "I say that the focus of teacher education should be on language and culture, rather than on ethnic studies, I mean on the interaction of the school with the family, home and community" (p. 2).

Diaz (1987) also acknowledges the importance of the cultural connection between home and school when he argues:

> In contrast to the past researchers have recently been focusing on how schools can capitalize on cultural practices by incorporating them into classroom activities and lessons. Such attempts to 'match' culture with educational activities are relatively new, and their effectiveness remains to be tested longitudinally. Still, increasing evidence points to their effectiveness in promoting academic achievement. (p. 9)

916

Cuevas (1980), drawing upon the research of Barnes (1977), argues that teachers need to be aware that they do not participate in or promote social behaviors that put students of color down or are culturally offensive, for example:

> Establishing and adhering to an etiquette of race relations in the classroom whereby the minority student is low person on the totem pole.
>
> Patting minority children on their heads in a condescending way
> Referring to minority students as "you all," "you people," "your kind. (p. 39)

Writing in a similar vein, Trueba (1983), after conducting an anthropological study in the Ocean View School District in California argues that some teachers are successful at coaching Mexican-American students because they are able to adopt strategies to comfort them. For example, Trueba points out that these teachers code-switch from English to Spanish and use appropriate touching behavior. Also, Mitchell (1985) observed one teacher's "effective use of language" in a black day care center. She concluded that because the teacher regularly switched back and forth between formal speech and informal speech that was used in the community, the students were better able to adjust to the traditional school's codes and were comfortable with curriculum content.

Quintanar-Sarellana (1991) administered a cultural awareness questionnaire to 71 teachers in bilingual programs and 56 teachers in English-only programs. She discovered that teachers, who work in a bilingual program perceive the language and culture of minority students more favorably. Quintanar-Sarellana (1991) argued that, the study points up two key elements for teacher training. The first one deals with the sociocultural knowledge of the teacher, "understanding of their own culture, as well as appreciation of other cultures and intercultural knowledge" (p. 21). The second one deals with, "the need to recruit and train Hispanics to be teachers" (p. 23).

These studies clearly suggest that teachers need to be aware, accept and affirm the culture and language their students bring to school. This acceptance and affirmation of the students' home culture and language is important to school success of LEP students. However, the lack of studies that pursue a particular chain of inquiry in this area suggest that much could remain a mystery about language, culture, and schooling for LEP students.

Multicultural Analysis

The general approach taken in most of these studies seems to be teaching the exceptional and culturally different. They point out that school/classroom teaching is adjusted to accommodate the needs of culturally different learners. For example, Cazdens and Mehan quote Berstein (1972), "If the culture of the teacher is to become part of the consciousness of the child, then the culture of the child must first be in the consciousness of the teacher." Similarly, Diaz's (1987) recognization of the importance between the students' culture and school activities for promoting learning is based upon instruction that builds bridges between the home and school in order to enable the student to catch up or fit in.

Cuevas' (1990) study also seems to support the teaching the exceptional and culturally different approach to multicultural education. Mitchell's (1985) sample is too small in sample size (one person) to speculate on the approach to multicultural education.

Trueba (1983), however, argues for an education that is multicultural and social reconstructionist. For example, Trueba (1983) posit:

> Teachers and administrators must come to the realization that the school is multi-ethnic and multicultural, that a pluralistic philosophy of education has implications for resource allocation and distribution of power at all levels, and that equity requires fairness, that is, no differential treatment of teacher, parents, and children on the basis of cultural or linguistic characteristics. ...Equity implies a measure of political equality, the sharing of power (decision-making especially) by all ethnic group involved in the school. (p. 412)

It is interesting that with the exception of a few researchers (e.g., Trueba), most of the discussions regarding culture and language have an implicit and often explicit message that LEP students should be assimilated into schools. There is rarely discourse or a plan of action regarding changing schools to better meet the needs of the LEP students. Also, assimilation into schools as they presently exist ignores structural and institutional bases of oppression.

To a great extent, the LEP students' language and culture is seen as a "problem" to be fixed by the school. In many ways, the term "limited" suggests a short fall, a minus, not a plus, and supports a deficit perspective when thinking about students who are non-native English speakers.

Community Participation

Based upon interviews with four different groups of bilingual teachers located in four different California schools, Ada (1986) pointed out that all groups agreed on the importance of home/community-school participation. She reports that one teacher suggested that "teacher education programs should include inservices from community leaders." A second bilingual teacher suggested that "teacher education programs should include a form of internship in community projects so that teachers might gain a holistic view of the community and become involved in wider societal issues" (p. 390).

Cuevas (1980) offers several recommendations for involving parents in school activities. The activities include home visits, using parents as resource persons, conducting parent group meetings, and tapping into community resources.

Bermudez & Padron (1988) reported on a collaborative effort between the University of Houston-Clear Lake and local school districts to develop a parent training program that included preservice and in-service teachers. The goal of the program (pertinent to this paper) was to help the teachers understand the cultural and linguistic barriers to school involvement that the parents of LEP students face. The results of the study were that teachers' attitudes about minority parent involvement in school were positively changed.

Moll and Diaz (1987) conducted two case studies with Hispanic working class students and their teachers and concluded that an understanding of the students' community and knowledge of the community's resources are important to the improvement of classroom instruction.

Walker (1989), in a study of Hmong culture, pointed out that Southeast Asian parents are interested in participating in their children's education. She states that, "Education is a family affair. The entire family may learn from a homework assignment" (p. 176).

Multicultural Analysis

The studies in the community section seem to promote community involvement in a human relations manner. The emphasis is on teachers learning the school community, eliminating any negative stereotypes about the students and their home life, and replacing them with feelings of acceptance and tolerance. Also, the emphasis is on helping parents develop positive feelings about the school. There is rarely any discussion concerning parents or community members becoming actively involved in the education decision-making process

Pedagogy

Cazden and Mehan (1989), Diaz, (1987), Mehan and Trujillo (1989) all posit the importance of a context specific view of human behavior.

Cazden and Mehan (1989) reviewed the three following studies: (1) Cazden (1972), who examined the average sentence length of two students' speech, one a middle class boy who was judged to be an excellent reader and the other a working class girl who was virtually a nonreader; (2) Heider, Cazden & Brown (1968) who examined the description (density of criteria attributes) of a picture of one animal from a large array by white middle class ten-year-old boys and white working class ten-year-old boys; (3) Diaz, Moll, & Mehan (1986) and Moll and Diaz (1987) who observed the same elementary students during reading lessons taught in Spanish and English. Based upon this review, Cazden and Mehan (1989) argue that the context of the task greatly influences student learning. Cazden and Mehan (1989) observe:

> This context-specific view of human behavior contributes to our understanding of the poor school performance of many low-income and linguistic minority students. Instead of blaming school failure on student characteristics that the school cannot change, teachers should reconsider aspects of the classroom environment that are within their control. Studies such as those we have reviewed here suggest the need for beginning teachers to vary instructional circumstances in order to take full advantage of students' often unrecognized resource. (p. 49)
>
> If students do not at first respond in ways that teachers hope and expect, teachers should not immediately assume that the students do not know or do not care. Instead, they should consider aspects of the classroom environment that might be changed. (p. 49)

At the 1989 Linguistic Minority Research Project Conference, Mehan and Trujillo drawing upon the findings of these and other studies claimed that "Intelligence is not a general, context-independent ability, it is a context-specific skill which varies from one type of situation to another" (p. 1). During the discussion period at the Conference, Mehan added, "If there is a single word that could summarize everything I have to say, it is context. The idea of context is a fundamental ingredient of the knowledge base for the beginning teacher, and the concept of intelligence demands a contextual analysis" (p. 2).

Garcia, Carter, Garcia, & Sevens (1989) conducted a study to determine the attributes of "effective" schools for linguistic minority students and discovered (pedagogically speaking) that instructional activities organized in a collaborative small heterogenous group setting worked best for LEP students. It was also important to limit individual instructional activities, such as worksheet and workbook work, as well as the use of competition as a motivational device.

Kagan (1985) argues that cooperative learning styles are important to the learning of linguistic minority students. However, he cautions that teachers must be careful because "language minority students are by no means exclusively oriented toward cooperative learning" (p. 26). However, during his keynote address at the 1987 University of California Linguistic Minority Research Project Conference, Kagan claimed that the results from four major national studies in which cooperative learning methods were studied revealed that, "Anglo students continue to gain at or above the levels they gain in traditional classes and the minority students show a large increase. There's an actual closing of the school achievement gap over time" (p. 4). Kagan also added that, the second major finding in cooperative learning has to do with improved ethnic relations among and between students (p. 4).

Cazden and Mehan (1989) discuss the concept of homogenous grouping and cooperative grouping as it relates to language minority students. They argue that the works of scholars in this area (e.g., Cohen, 1986; Kagan, 1986; Oakes, 1985; Slavin, 1983) point up that homogenous grouping does not successfully aid the academic success of language minority students, and because of this beginning teachers need to consider alternatives.

> Cooperative learning, the structuring of classrooms so that students work together in small interdependent teams, and heterogenous grouping, whereby more sophisticated learners are placed with less sophisticated learners, are two alternatives that may bring about educational outcomes that are more positive than those presently provided by homogenous ability grouping. (p. 53)

Berg (1987) makes a similar observation, "...teachers need not have a specific curriculum or teaching style for each cultural group. ...a teacher needs to have a wide variety of accessible teaching strategies to draw from based on the students' needs" (p. 18).

Along with an understanding of context-specific instruction and cooperative grouping studies, some pedagogical attention has been given to Berg's (1987) proposal. Berg (1987) argues for instructional strategies that allow cultural differences to emerge naturally in the classroom. Somewhat related, Cazen and Mehan (1989) argue for making certain that LEP students understand classroom rules and norms. For example, Cadzen and Mehan believe that students' knowledge of classroom rules and norms is positively correlated with school success.

Trueba (1988) in a study to discover the instructional difficulties faced by teachers and to identify successful instructional strategies for LEP students, argued that "the literacy problem faced by linguistic minorities is deeply related to their lack of such cultural knowledge that is presumed by the instructors and writers of textbook material" (p. 356). He adds that, "effective instruction for linguistic minority children in cultural transition, even if it must be conducted in English, a language not well understood by these children, can still be tailored to children's cultural knowledge and experience" (p. 358). He suggests that teachers of LEP students need to experiment with different instructional settings, strategies, and experiences.

Short and Spanos (1989) conducted a study on content-based instruction, mathematics, with LEP students. The study involved collaborative research with mathematics educators at several two-year colleges with a high enrollment of LEP students. The study's intervention was a set of materials designed to be used as a language focused supplement for beginning algebra classes. The researchers discovered that both the language minority students and the majority students had difficulty doing problem-solving activities because of their lack of proficiency in the language of mathematics. One major implication for teacher training, suggested by this study, is to provide workshops and seminars so content teachers can be more informed about how to include language objectives and increased communication in their classes.

Ada, (1986) after an interview with thirty-eight bilingual teachers regarding the classroom problems they face and how teacher education programs might better address these problem, argues that teacher training programs for LEP students need to teach them empowerment skills. She posits that, "many teacher education programs seem designed to train teachers to accept social realities rather than to question them" (p. 388). Ada (1986) points out that teacher education programs need to teach the future teacher the importance of peer support. Students need the opportunity to live, study, and possibly teach in a country where the language they will be teaching is spoken, and need to better integrate theory and practice. Ada (1986) noted that the strongest criticism of teacher education programs was that the faculty in the school of education did not teach the way they argued that teaching should take place.

Aronson (1985) argues that the overemphasis on classroom competition has inhibited the achievement of LEP students. He reminded educators that Mexican-American students perform the most effectively in learning settings that promote cooperative efforts that are in pursuit of common goals. Kegan (1985) speaking at the same Linguistic Minority conference supported Aronson's views but added:

> ...language minority students are by no means exclusively oriented toward cooperative learning. It is true that they tend to prefer cooperation over competitiveness, and that in the usually competitive framework of North American classrooms, this cul-

> tural preference affects their educational achievement. Yet it is essential that students adapt to both styles of learning. No one style should be exclusively accepted as "correct." Students must learn to discriminate which style is appropriate for what contest. (p. 26)

Walker (1989) in a study of the Hmong students in school argues that most of the in-service training for teachers about Hmong have been developed in isolation, without information gained being shared among teachers.

Multicultural Analysis

The importance of context-specific instruction and the importance of using grouping (mostly cooperative groups) were the two major areas of focus in this section. These studies for the most part contain discussions of the use of these pedagogical strategies in teaching the exceptional and cultural different manner, with some attention to human relations. This means that the discussion of context is mostly in relation to modification of the teaching environment and acknowledges and accepts the culture and language differences the students bring to school. Similarly, the discussion of grouping suggests cooperative grouping as a pedagogical strategy to facilitate the school work of Hispanic students, because it is believed that by having students work together student achievement will be enhanced.

Similarly, Garcia, Carter, Garcia, & Stevens (1989) argue, "Effectiveness is the result of cooperative and collaborative endeavors of staff, administration, and community." And, "The effective school is outcome focused, not input focused. Like industry it constantly improves the quality of its `product'." Additionally, the way to promote classroom instruction for LEP students, suggested by Aronson (1985) and Kagan (1985), seemed to be "cooperative learning." Both concepts, collaboration and cooperative learning are important and fundamental to the Human Relations approach and serve to identify this approach, especially when little or no discussion related to empowerment, social stratification, and institutional discrimination is included.

The ideas proposed by Ada (1986) in preparing teachers to work with LEP students are in keeping with the education that has a multicultural and social reconstructionist approach on the Grant and Sleeter paradigm. Ada posits:

> I believe the views of Freire (1982a, 1982b) and Giroux (1985) are correct: schools do hold out the possibility of critical analysis and reconstruction of social reality through meaningful dialogue be-

> tween teachers and students, by a process termed "transformative education." (p. 387)

The Short & Spanos (1989) study is designed to inform teachers about how to work more effectively with the Exceptional and Culturally Different. However, it does not argue for instructional strategies that will teach the students to question why they are considered "limited" English proficient, instead of students acquiring and enriching speaking and writing excellence in two languages.

Assessment

McLean's (1981) findings from the first national assessment which included determining the scope of training of teachers and the teacher competencies needed for working with LEP handicapped students revealed the following as important: a desire to work with LEP handicapped students; a sensitivity and knowledge about working with LEP students; the knowledge and skills necessary for relating to the parents of LEP handicapped students; the knowledge, skills, and methods for teaching LEP handicapped students; and the ability to develop curriculum and instructional plans to meet their needs.

Baca, Fradd and Collier (1990) reported a follow-up of the McLean (1981) study conducted in three states, California, Colorado, and Florida. Results important to this paper from the California study, (Baca, 1987) that surveyed 420 special education/bilingual educators and administrators in attendance at a conference on LEP handicapped revealed the following:

> 58 percent of the participants reported that the colleges and universities in their area were training bilingual special education personnel, 20 said no, and 22 reported they didn't know.
>
> The participants ranked the competency for dealing with knowledge of legal issues regarding minority students as the most important.

The Cross Cultural Special Education Network (1987) surveyed 150 school districts in Colorado regarding bilingual special education. Responses from 114 school districts revealed the following competencies as necessary or important for working with LEP students:

> ...knowledge and sensitivity toward the history and culture of LEP students, ability to work with an interpreter in assessment and instruction, knowledge of different cultural perception of handicapping conditions, knowledge of tests and technique for evaluating the mental capabilities of LEP students, knowledge of

> general instructional methods applicable to LEP handicapped children, the capacity to integrate teaching techniques from the field of bilingual education and special education, the knowledge of methods technique for developing material especially for LEP handicapped children, and the knowledge of methods for dealing with the parents of LEP handicapped children. (p. 11)
>
> The following were reported to be significant; knowledge of the educational implications of social class background and the process of acculturation, knowledge of test and techniques for evaluating language dominance and proficiency versus language disability, and knowledge of the legal issues concerning the education of LEP students. (p. 11-12)

Special education directors and ESOL supervisors in the 60 Florida school districts with identified LEP students received copies of the questionnaire used in California and Colorado. Fifty-nine of the school districts responded, with results similar to Florida.

Based upon their surveys, Baca, Fradd, and Collier (1990) recommended that, "preservice and inservice education be given high priority and be made available both by school districts and universities." They also suggested that awareness training for special education personnel and administrators be increased in all states highly affected by the presence of LEP students (p. 11).

In another study designed to identify the competencies needed by LEP handicapped students Fradd, Algozzine, & Salend (1988) had 51 respondents from New York and 51 respondents from Florida complete a competency survey. The respondents were grouped into three areas: teachers of bilingual education, teachers of special education and teachers of bilingual special education. The survey included 15 general competencies identified in a review of the literature which were assumed important to personnel engaged in special education teaching in bilingual education. These competencies were in the areas of testing, human growth and development, characteristics of handicapped students, budgeting, culture, resource utilization, proficiency in both English and another language, linguistic analysis, use of research information, interpersonal skills, parent involvement, moving students from non-English into English, and materials development. All three groups of teachers ranked all the competencies in each of the areas listed above as being fairly important. However, all three groups saw competency in moving students out of non-English language and into English as extremely important.

Multicultural Analysis

Most of the studies in this section have to do with the identification of competencies for working with LEP handicapped students.

The type of competencies identified (e.g., sensitivity, knowledge of different cultural perceptions of handicapped, Cross Cultural Network, 1987) are more closely associated with teaching the exceptional and culturally different approach. These instructional competencies, for the most part, are designed to move LEP students into the mainstream, often at the expense of the students' native language. Other competencies tend to be associated with the Human Relations approach, for example, to promote good feeling between the home and school.

Review Discussion

Preservice, Subjects, and Nature of Studies. Research studies on preservice teacher preparation programs for LEP students are few. In fact, most of the studies located for this paper were done mainly with experienced teachers. However, some of these studies (e.g., Cazen & Mehan, 1990) did suggest implications for beginning teachers. From this it could be reasoned that teacher preparation programs for LEP students need to make certain that their students leave the university understanding and affirming the importance of: (1) home culture and language of the students they teach; (2) students' home and community participation in school and classrooms activities; (3) the inter and intra relationship of instruction and context; and (4) cooperative learning.

Because the research base on preparing teachers to work with LEP students is so limited, it argues for a major research thrust in the following areas:

- **In-service training**
- **Research techniques**
- **Competencies in training LEP handicapped students**

Most of the studies reviewed in this section were aimed at positing what teachers need to know, (mainly about the students) in order to successfully teach LEP students. The studies (e.g., Cuevas, 1980) argue that a fundamental awareness of students' cultural history, which is grounded in respect and takes into account cultural "no-no's," for example, patting the head of a LEP student are important to instructional success. Also, these studies (e.g., Moll & Diaz, 1987) argue that teachers' understanding of the school community and how to involve parents and other community members in the school's program is vital. Besides, knowing about the students and their community, several researchers (e.g., Garcia, Carter, Garcia, & Sevens; Aronson, 1985) identify cooperative groups, and a de-emphasis on classroom competition as important to classroom success for LEP students. Similarity, researchers (Trueba, 1988; Short & Spanos, 1989) pointed out teachers must understand that there are other important factors besides proficiency in English. For example, LEP students may lack cultural knowledge about schooling, e.g., the language of mathematics. In addition, textbook usage procedures need to be addressed simultaneously with the goals of English proficiency.

The research techniques employed in many of these studies are anthropological, including the use of questionnaires, interviews, and observations. Most of the studies seem to have been conducted within a short time frame and to be singular in occurrence. Several of the researchers seemed to be concerned about similar issues, for example, cooperative grouping. However, there were few, if any, studies that replicated previous studies.

Several studies (Baca, Fradd, & Collier, 1990) sought to identify the personal competencies that teachers working with LEP disabled students need to have. The competencies are very similar to those identified for teachers working with regular LEP students. That is, knowledge and sensitivity regarding LEP handicapped students, understanding of their home life, and having the ability to work with their parents, and skills in moving students from non-English speaking to English proficiency. This set of studies seems to have a more central focus and the researchers seem to be drawing upon the work of one another. Fradd, for example, has conducted surveys with researchers in several states.

General Discussion

This essay started by reminding the reader that research in teacher education is thin, and that research both at the preservice and in-service level for preparing teachers to work with LEP students would be especially thin. This is so. A number of these studies, complete with narrative and references are difficult to locate through the normal retrieval process, i.e., through ERIC or a journal publication search. However, often available are short synopses of the results of studies, without research design, population sample, and other important information needed for replication or evaluation. The more coherent research on teacher preparation for LEP students seems to come from those working with teacher training of LEP disabled students. However, these one-time research findings seem to come solely from survey data collection, rather than longitudinal studies employing a variety of data collection methods. Nevertheless, there is a growing body of literature discussing the needs of LEP teachers, and from this literature a pattern of instructional practice important to LEP teachers is emerging.

Programmatic Patterns and Recommendations

From analyzing the research literature on the preparation of teachers (both preservice and in-service) to teach LEP regular and disabled students it can be reasoned that there are some recommended "best practices" that should be a part of every teacher preparation. These are:

Teachers must develop a cultural sensitivity and awareness, beginning with their own culture, that will allow them to work with students from any culture in a manner that shows awareness, acceptance/appreciation and affirmation of the culture.

Teachers in preservice and in-service programs must learn the importance of knowing and understanding the home and community life of their students. They must be prepared with the anthropological and sociological tools so they explore and learn about their students' lives in a way that informs without offending their students.

Teachers must developed skills in using grouping techniques and patterns that foster the learning styles of their students. Cooperative groupings and other small heterogenous arrangements seem to promote the social and academic success of LEP students; however, teachers need to know and understand the dynamics that can occur when groups are formed.

Teachers need to understand the importance of "context" in the instructional process. How (e.g., related to the students background) an educational concept situated in the teaching process influences students' level of understanding.

Teachers need to determine the approach to multicultural education they wish to adhere to: in promoting cultural awareness, in pedagogical instruction, in community/home-school involvement, and in educational assessment.

Reflections and Direction for Future Research

One decade ago, September 1981, Chamot (1981) in an article "Applications of Second Language Acquisition Research to the Bilingual Classroom," after reviewing the educational literature regarding teaching LEP students, identified four areas of research that should be applied to teaching LEP students: (1) similarity of first language teaching to second language teaching; (2) social, affective and cognitive factors; (3) second language input; and (4) second language learning in school settings (p. 1). Chamot (1981) further identified sub-topic areas under each of the topic areas. What is of interest to this paper is to what extent these four areas and the sub-topic points were integrated into teacher education programs for preparing teachers for LEP students. Also of interest were these topic areas and the sub-topic areas included in the research on LEP teacher education programs. Chamot (1981) elaborates on her first topic area as follows:

Topic 1

> Because second language learning is similar to first language learning, teachers should:
>
> Expect errors and consider them as indicators of progress through stages of language acquisition.
>
> Respond to the intended meanings children try to communicate.
>
> Provide context and action-oriented activities to clarify meanings and functions of the new language.
>
> Begin with extensive listening practice, and wait for children to speak when they are ready.
>
> Avoid repetitive drills and use repetition only as it occurs naturally in songs, poetry, games, stories and rhymes. (p. 6)

The review of research literature for this essay reveals that for Topic 1 the sub-topic area "provide context" was examined and discussed, the other sub-topic areas received little or no mention in the research literature.

For her second topic areas, Chamot (1981) argues:

Topic 2

> Because social and affective factors and differences in cognitive learning styles influence second language learning, teachers should:
>
> Foster positive, caring attitudes between limited- and native-English-speaking children.
>
> Plan for small-group and paired activities to lessen anxiety and promote cooperation among all children.
>
> Provide for social interaction with English-speaking peers.
>
> Vary methodology, materials, and types of evaluation to suit different learning styles.

> Build understanding and acceptance of cultural diversity by discussing values, customs, and individual worth. (p. 7)

The review of research literature for this essay reveals that for Topic 2 the sub-topic "cooperative grouping" and "appreciating the student's home culture" were examined, but the others received little or no attention.

Chamot's (1981) third topic area argues:

Topic 3

> Because the appropriate type of input is necessary for second language acquisition to take place, teachers should:
>
> Ensure that they model language that is meaningful, natural, useful, and relevant to children.
>
> Provide language input that is a little beyond children's current proficiency level but can still be understood by them.
>
> Plan for a variety of input from different people, so that children learn to understand both formal and informal speech, different speech functions, and individual differences in style and register. (p. 7).

The review of the research literature for this essay reveals that for Topic 3, I did not locate any research on teacher preparation programs that explicitly dealt with any of the sub-topics.

For her fourth topic area, Chamot (1981) argues:

Topic 4

> Because communicative competence in a second language does not provide children with sufficient skills to study successfully through the medium of that language, teachers should:
>
> Develop children's concepts and subject matter knowledge in their stronger language during the second language acquisition process so that they will be able to transfer these concepts to the new language.
>
> Use the second language for subject matter instruction when children reach the linguistic threshold needed to attach new labels to known concepts.

> Initiate subject matter instruction in the second language in linguistically less demanding subjects, such as math.
>
> Emphasize reading and writing activities in the second language as soon as children are literate in the first language.
>
> Realize that tests of communicative competence evaluate children's ability to function in social setting, not their ability to perform successfully in academic settings. (p. 7)

The review of the research literature for this essay reveals that, for topic area 4, there was no research on teacher preparation programs preparing students to teach LEP students that explicitly dealt with any sub-topics.

What did we learn from this examination? 1. Houston, Haberman, and Sikula's observation that "Although the importance of research is espoused, little progress is being made." (p. ix) seems to be accurate. 2. Teacher preparation programs do not see results from this research as serving to influence their research agenda or they are not interested in using this research. 3. Research reports are not readily available to teacher educators preparing teachers to teach LEP students.

Beyond Behaviorist Conceptions of Knowledge

Much of the research focused on changing teachers' beliefs (the home and culture of LEP students is acceptable) and behaviors (move to context specific instruction, use more cooperative grouping). As Grant and Secada (1990) observed, this is not surprising in view of the large bodies of research on teacher expectancies (Dusek, 1985). Nevertheless, it is important that research go beyond concepts of changing teacher beliefs and behaviors about working with diverse students. It is also important to understand how these teacher beliefs and behaviors impact on classroom management and instructional preparation. How biases toward some students and/or incorrect information about students can be greatly reduced or eliminated in teacher education programs. Additionally, it is important to learn how stereotyped and biased student expectations might be replaced with more direct methods of accessing students abilities (Grant & Secada, 1990).

It is important that research examine the schools' goal of knowledge utilization. Much rhetoric is given to students obtaining knowledge so they can think critically. Critical thinking is important, but equally as important is what the critical thinking is about. Is the school's goal of knowledge utilization for LEP students mainly to help them fit into society as it exists and thereby give up their culture, or is it to learn how to keep their culture and change society to the better?

Conclusion

There is much to learn about preparing teachers to teach LEP students. Research should play a major role in giving directions to what teacher educators include in their programs and to what teachers do in the classroom. Presently, however, the quantity and quality of this research isn't available. It should be, because until we completely understand how to educate LEP students we put at risk their life chances and opportunities.

References

Ada, A. F. (1986). Creative education for bilingual teachers. Harvard Educational Review, 56(4), 386-393.

Baca, L., Fradd, S. H., & Collier, C. (1990). Progress in preparing personnel to meet the needs of handicapped limited English proficient students: Results of a survey in three highly impacted states. The Journal of Educational Issues of Language Minority Students, 7, 5-20.

Berg, M. (1987). Matching instructional strategies with purposes for diverse settings. In R. Dash (Ed.), Roundtable: The challenge -- Preparing teachers for diverse student populations (pp. 15-20). Northridge, CA: Far West Laboratory.

Bermudez, A. B., Fradd, S. H., Haulman, A., & Weismantel, M. J. (1989). Developing a coordination model for programs preparing personnel to work with LEP students. The Journal of Educational Issues of Language Minority Students, 5, 79-95.

Cazden, C. B., & Mehan, H. (1989). Principles from sociology and anthropology: Context, code, classroom, and culture. In M. C. Reynolds (Ed.), Knowledge base for the beginning teacher (pp. 47-57). Oxford, England: Pergamon Press.

Chamot, A. U. (1981). Applications of second language acquisition research to the bilingual classroom. (National Clearinghouse for Bilingual Education)

Cohen, E. G. (1986). Designing group work-strategies for the heterogeneous classroom. New York: Teachers College Press.

Cuevas, G. (1980). Staff development. In L. A. Valverde, R. Castro Feinberg, & E. M. Marquez (Eds.), Educating English-speaking Hispanics (pp. 37-50). Alexandria, VA: Association for Supervision and Curriculum Development.

Cummins, J. (1986). Empowering minority students: A framework for intervention. Harvard Educational Review, 56(1), 18-36.

Diaz, E. (1987). Cultural patterns affecting teacher and student expectations. In R. Dash (Ed.), Roundtable: The challenge -- Preparing teachers for diverse student populations (pp. 9-13). Northridge, CA: Far West Laboratory.

Dusek, J. B. (Ed.). (1985). Teacher expectancies. Hillsdale, NJ: Lawrence Erlbaum.

Fradd, S. H., Algozzine, B., & Salend, S. J. A survey of training needs in bilingual special education. (1988). The Journal of Educational Issues of Language Minority Students, 2, 5-16.

Freire, P. (1982a). Education for critical consciousness. New York: Continuum.

Freire, P. (1982b). Pedagogy of the oppressed. New York: Continuum.

Garcia, E. E. (1990). Educating teachers for language minority students. In W. R. Houston (Ed.), Handbook of research on teacher education (pp. 717-729). New York: Macmillan.

Garcia, E. E., Carter, T., Garcia, E., & Stevens, K. (1989). In K. Hakuta (Ed.), University of California Linguistic Minority Research Project Research and Policy Series (No. 3). Santa Barbara: University of California, LMRP.

Garza, S. A., & Barnes, C. P. (1989). Competencies for bilingual multicultural teachers. The Journal of Educational Issues of Language Minority Students, 5, 1-25.

Giroux, H. A. (1986). Radical pedagogy and the politics of student voice. Unpublished manuscript.

Grant, C. A., & Secada, W. G. (1990). Preparing teachers for diversity. In W. R. Houston (Ed.), Handbook of research on teacher education (pp. 403-422). New York: Macmillan.

Grant, C. A., & Sleeter, C. E. (1985). The literature on multicultural education: Review and analysis. Educational Review, 37(2), 97-118.

Grant, C. A., Sleeter, C. E., & Anderson, J. E. (1986). The literature on multicultural education: Review and analysis. Educational Studies, 12(1), 47-71.

Hamayan, E. V. (1990). Preparing mainstream classroom teachers to teach potentially English proficient students. Proceedings of the First Research Symposium on Limited English Proficient Students' Issues (pp. 1-22).

Houston, W. R., Haberman, M., & Sikula, J. (1990). Preface. In W. R. Houston (Ed.), Handbook of research on teacher education (pp. ix-xi). New York: Macmillan.

Kagan, S. (1985). Cooperative learning and schooling outcomes for socio-cultural linguistic minority students [Summary]. In H. Mehan, Jr., H. T. Trueba, & C. Underwood (Eds.), Schooling Language Minority Youth: Proceedings of the University of California Linguistic Minority Project Conference (Vol. 1) (pp. 25-27).

Kagan, S. (1986). Cooperative learning and sociocultural factors in schooling. In Beyond language: Social and cultural factors in schooling language minority students (pp. 231-298). Los Angeles: California State University, Evaluation, Dissemination, and Assessment Center.

Kagan, S. (1987). Cooperative learning for linguistic minority students [Keynote Address]. Schooling Language Minority Youth: Proceedings of the University of California Linguistic Minority Research Project Conference (Vol. 3) (pp. 3-13).

Mehan, H., & Trujillo, T. (1989). Teacher education issues. In K. Hakuta (Ed.), University of California Linguistic Minority Research Project Research and Policy Series (No. 4). Santa Barbara: University of California, LMRP.

Mitchell, J., & Rice, N. (1985). Teacher strategies for mediating social control and instructional practices [Summary]. In H. Mehan, Jr., H. T. Trueba, & C. Underwood (Eds.), Schooling Language Minority Youth: Proceedings of the University of California Linguistic Minority Project Conference (Vol. 1) (pp. 22-24).

Moll, L., Diaz, E., & Anderson, A. (1985). Community and family influences on literacy development [Summary]. In H. Mehan, Jr., H. T. Trueba, & C. Underwood (Eds.), Schooling Language Minority Youth: Proceedings of the University of California Linguistic Minority Project Conference (Vol. 1) (pp. 19-21).

Moll, L. C., & Diaz, S. (1987). Change as the goal of educational research. Anthropology and Education Quarterly, 18, 300-311.

Oakes, J. (1985). Keeping track. New Haven, CT: Yale University Press.

Quintanar-Sarellana, R. (1991, May). Training teachers for a multicultural society: Teachers' perceptions of the language and culture of linguistic minority students. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Rosado, L. A., & Aaron, E. B. (1991). Parental involvement: Addressing the educational needs of Hispanic inner-city parents. The Journal of Educational Issues of Language Minority Students, 8, 23-29.

Short, D. J., & Spanos, G. (1989). Teaching mathematics to limited English proficient students. Washington, DC: Center for Applied Linguistics. (ERIC Clearinghouse on Languages and Linguistics)

Slavin, R. E. (1983). Cooperative learning. New York: Longman.

Sleeter, C. E., & Grant, C. A. (1987). An analysis of multicultural education in the United States. Harvard Educational Review, 57(4), 421-444.

Sleeter, C. E., & Grant, C. A. (1988). Making choices for multicultural education: Five approaches to race, class, and gender. New York: Merrill.

Trueba, H. T. (1983). Adjustment problems of Mexican and Mexican-American students: An anthropological study. Learning Disability Quarterly, 6, 395-415.

Trueba, H. T. (1988). Instructional effectiveness: English-only for speakers of other languages? Education and Urban Society, 20(4), 341-362.

Ventriglia, L. A functional approach to mainstreaming limited-English-proficient students. (National Clearinghouse for Bilingual Education)

Walker, W. D. (1989). The challenges of the Hmong culture: A study of teacher, counselor and administrator training in a time of changing demographics (Excerpts from). Unpublished doctoral dissertation, Harvard University.

Watkins, A. (1987). Effective teachers of diverse students. In R. Dash (Ed.), Roundtable: The challenge -- Preparing teachers for diverse student populations (pp. 21-26). Northridge, CA: Far West Laboratory.

Zeichner, K. (1991, April). Teacher education for social responsibility: The conception of teaching expertise underlying elementary teacher education at the University of Wisconsin - Madison. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Response to Carl Grant's Presentation

Margarita Calderon
University of Texas, El-Paso

Quality teacher preparation is the most worthy goal for ensuring success of limited English proficient students. It is clear that the LEP students' schooling process is threatened unless immediate and aggressive efforts are undertaken to attract, prepare, and retain teachers who are well prepared to meet their needs.

I would like to organize my response to the Grant paper through the following framework: (1) attracting teachers of LEP students (**recruiting);** (2) supporting the teacher preparation phase **(preparing);** (3) assisting teachers of LEP students in the first years of teaching **(inducting);** and (4) beyond the first years: retaining and upgrading the skills of teachers of LEP students **(retaining or staff development)**. I will cite research in each area and make connections to the Grant paper while extending the discussion to teacher support systems needed at a more macro level -- the organization structures of universities and schools, and a micro level -- the processes of training and coaching teachers of LEP students. I will end by sharing an innovation in pedagogy and the use of cooperative learning for teacher training.

Attracting and Recruiting Teachers of LEP Student

From a recent body of research in Texas and California (Cuellar & Huling-Austin, 1991; Tomas Rivera Center, 1990, 1991) we find that:

1. Minority students need primary language role models.
2. Minority teachers bring additional insights and perspectives to the job of teaching.
3. All students benefit from having teachers who represent today's cultural society.
4. An ethnically-diverse teaching force can bring stability to the staffing of schools in regions that have traditionally experienced high teacher turnover rates.

In light of Grant's review, many negative effects of not having a minority or ethnically-diverse teaching force are self-evident. While it is important that "teachers develop a cultural sensitivity and awareness that will allow them to work with students from any culture in a manner that shows awareness, acceptance/appreciation and affirmation of the culture" (p.24), universities and schools must have structures that facilitate formal opportunities for recruiting ethnic and cultural representation in teaching education and the teaching profession. Thus, the area of recruiting also needs to be included into the realm of effective teacher preparation. Without teachers, our efforts to reform schools and restructure education will count for nothing.

1. Recruiting

From 1988 to 1991, the Texas Education Agency funded three cycles of grants focused on attracting and retaining minority/bilingual teachers in the teaching profession. These projects included research components that looked at "most promising practices" in the area of recruiting and preparing bilingual and monolingual teachers of LEP students. Published reports are now available from the Texas Education Agency as well as the Journal of Teacher Education, which devoted a volume to the results of these studies.

Specific guidelines are delineated by W.R. Houston and M. Calderon (1991) in these publication for university personnel, public school educators, state legislators, state educational agency personnel, educational organizations, community and business groups, and researchers.

2. Preparing Teachers of LEP students

In 1988, The Tomas River Center (TRC) identified forty-six institutions of higher education in the Southwest that enrolled significant numbers of Latinos in their teacher training programs. The TRC researchers found that there were various forms of recruitment and retention efforts, but no teacher-training programs integrated the full range of effective practices. This led the Tomas Rivera Center to secure a grant from the Exxon Educational Foundation to create four research and development projects to increase the supply of well-prepared minority teachers who will teach minority students.

Four universities are currently being funded to research and develop comprehensive programs for this purpose (San Diego State University, San Bernardino State University, Southwest Texas State University, The University of Texas at El Paso). Programs must incorporate a set of interrelated practices in campus-wide efforts. They must, in short create learning communities that integrate recruiting, student advising, basic skills development, appropriate content for working with language minority students, and supportive environments that enable student to maximize their performance.

All four universities embraced the concept of cooperative learning and are building cooperative learning communities in a variety of ways at the university level. These include offering specialized teacher mentoring and tutoring services, setting up "buddy system", developing college success skills in groups, and learning how to become teachers and mentors in a collaborative content of school reform.

3. Inducting Teachers of LEP Students

From 1989 to 1991 the Texas Education Agency awarded eight research grants for "Enhancing the Quality and Retention of Minority Teachers and Teachers in Critical Shortage Areas (when included non-bilingual teachers of LEP students)." As the nation's third largest state in minority population, and faced with a shrinking number of teachers, who are constantly blamed for the failures of their students, Texas began to look at other states' education programs and decided to focus the research on minority issues.

These projects set out to build a support network for the first year minority/bilingual teachers, to motivate teachers to stay in the profession; to enhance their knowledge and skills regarding language minority instruction; and to improve their content base in critical shortage areas (e.d. science and math en espanol). The common element in these projects was that an experience teacher was well trained to coach the beginning teacher, while strengthening his/her own self-concept as a professional. In the process of peer-coaching, experienced teachers updated their knowledge and skills for working with language minority students (Ramirez, 1991). Overwhelmingly positive results in terms of retention, teacher satisfaction, teacher appraisals, and classroom instructional practices are documented in a publication soon to be released by the Texas Education Agency and the Intercultural Development Research Association.

4. Inservice/Staff Development of Teachers of LEP Students

In 1985 Secretary of Education Bell funded a study to look at a staff development model for training bilingual and monolingual teachers of LEP students. This was a continuation of a two year study that had been conducted in 5 school districts in Southern California. The Dept. of Education study look at the implementation of the Multidistrict Trainer of Trainers Institute (MTTI) that were implemented throughout the state of California and operating out of the County Offices of Education. This study focused on the content that teachers needed in order to shift into a constructivist approach to classroom instruction of LEP students, and also on the process of training and building support systems for teachers trying to shift into a new instructional philosophy and delivery system.

The results were well documented in a publication to OBEMLA (1986), and follow-up results were published in the NABE Journal (1988). In essence, the findings confirmed that although the content of the teacher training sessions is important, (1) the process for training and (2) follow-up support systems for collegial learning are critical. Without certain processes for preparing teachers, the content never transfers into their active teacher repertoire. Therefore, the teaching philosophies and teaching methods we would like teaches to espouse, never transfer into the classroom.

The elements of processes that help teaches transfer desired knowledge, behaviors and decisions into the classroom have been empirically tested for the past ten years (Calderon, 1981, 1982, 1984, 1986, 1991). These same elements of the wide-scale study of staff development practices were observed in the Texas study on the induction year of beginning minority teachers and their mentors.

We now predict, that these elements will also be essential in the undergraduate preparation of LEP teachers.

Briefly, these elements are: (1) presentation of theory, philosophy, research on each content area, followed by (2) extensive modeling of the teaching strategies, (3) analysis and discussion of student adaptation and modification to meet diverse needs, (4) extensive observation and practice in both simulated and real environments, (5) guided practice with feedback-peer-coaching, mentoring, video taping, (6) adaptation to curriculum and lesson planning, (7) reflection activities that lead to analysis of own teaching performance and decisions, and (8) self-directed collaborative study groups where colleagues continue to refine their practice.

More and more district and school level teacher development practices are beginning to incorporate all these elements into their staff development programs. A five-year study under the auspices of the National Center for Research on Effective Schooling for Disadvantaged Student (Calderon, Hertz-Lazarowitz, Tinajero, Duran, and Slavin) has been looking at teacher development through control and experimental classrooms of bilingual teachers. It has also studied a variety of ways of orchestrating staff development programs for ESL and bilingual teachers that incorporate these elements. Results so far, identify stages that teachers go through when attempting to implement student centered teaching innovations such as cooperative learning.

Staff Development Systems Approaches

Some examples currently being studies:

1. Single School + Researchers

- comprehensive staff development program (LEP instruction, peer-coaching, etc.)
- principal actively participates in staff development program, in the in-service session, coaches teacher support systems (Kauai Intermediate & High School; Waikiki and Liholiho schools).

2. Single School + District Bilingual Office = university Title VII fellowships + cadre of teachers as trainers of other teachers + researchers

- long term comprehensive staff development through collaborating agencies (San Antonio ISD).

3. District Deputy Superintendent of Instruction = central administrators + principals + volunteer teachers from 5 schools + researchers (Windward District in Oahu).

Innovations in Pedagogy

Grant (1991, p.24) identifies "contextualization" and "cooperative grouping and other small heterogeneous arrangements seem to promote the social and academic success of LEP students." However, he also cautions that "teachers need to know and understand the dynamics that can occur when groups are formed." This precaution should not be taken lightly. One of the biggest hurdles teachers have to overcome in effectively implementing cooperative learning is classroom management. Particularly when LEP students are involved. Figure 1 depicts the types of problems teachers have at each stage of implementation. Usually, the students' primary language is put on the back burner in order to facilitate the teacher's comfort with both the testing and academic demands of the school and the students' new role with cooperative learning.

Rachel Hertz-Lazarowitz is currently observing teachers of LEP students conduct cooperative learning. She uses a "Six mirror" instrument to observe:

1. the physical organization of the classroom (the types of learning groups);

2. the learning task (whether unitary, in pairs, groups with different structures of division and combinations;

3. the teacher behavior -- styles of instructional leadership from centralized to decentralized where decision making processes are distributed among groups of students.

4. the teacher behavior -- communication patterns with and among students.

5. the student social behavior -- from an isolated individual to an integrated member of the team.

6. the student academic behavior -- ranges from passive skills such as interacting only with the textbook and/or teacher, to highly complex, evaluation and creative skills synthesizing several sources of information with an interactive context (See Figure 2).

Researchers and teachers analyze these data and use it to integrate student background with cooperative learning strategies and thus contextualize learning. These observations have helped the researchers identify when changes in cooperative classrooms are only **superficial** and when they are **truly meaningful** and **constructivist** in nature. The observational tool, is also helping teachers learn to reflect about their teaching and identify areas for improvement when they meet in their study groups.

The Hertz-Lazarowitz study is part of a five-year longitudinal study being conducted by a team of interdisciplinary researchers from Johns Hopkins, UC Santa Barbara, Haifa University and the University of Texas at El Paso -- which is studying the effects of cooperative learning on LEP students in various sites. Several annual reports are currently available and several journal articles and book chapters describe student performance, development of literacy in two languages, the use of dynamic assessment, teacher development, the staff development and peer-coaching component, and the restructuring of school factors that are needed in order to effectively implement cooperative learning.

This longitudinal study is also a study of change: how students, teachers, administrators, researchers, schools, and school districts move progressively through stages of cooperation (or collaboration as it is termed at the school faculty level), and collective reflection, in an effort to implement programs that come closer to addressing the needs of LEP students.

Cooperative Learning as a Professional Development Tool

This use of cooperative learning as an implementation tool for building learning communities of teachers has also been studied through the MTTI studies and more recently through the induction programs. The elements of academic achievement, self-esteem, social skills and collaboration are discussed in relationship to adults participating in the Minority/Critical Shortage Beginning Teacher Project (Calderon, 1990). Among other findings, the researchers saw how cooperative learning structures helped teachers develop in several ways.

a. **Cooperative Learning for Academic/Instructional Development.** Cooperative Learning (CL) was used as the process for in-service training with four purposes in mind. (1) to teach the content requirements of the project, (2) to teach, apply and internalize principles of adult learning, coaching, feedback and support techniques, and (3) to conduct reflection, decision-making and problem-solving activities, and (4) to learn, vicariously, how to further enhance their use of CL strategies in their classroom.

b. **Cooperative Learning for Developing Collaborative and Social Skills.** It is a well known fact that teachers who have reached a high level of success as classroom teachers are the ones most likely to be selected as mentors or support teachers. However, expert classroom teachers may or may not be expert mentors or coaches of other teachers. The art of mentoring and/or peer-coaching requires certain social and collaborative skills. Yet, collaborative skills are not developed in isolation. If teachers have "grown professionally" in isolation for many years, the tasks and skills of working with peers need to be reviewed or developed. In order to foster an environment of trust and skills for coaching in their project, CL strategies were used where partners worked and learned together at the workshops through activities deliberately created to build trust, joint experimentation and appreciation for one another's talent.

c. **Cooperative Learning for Self-Esteem.** Typical staff development programs are sometimes so laden with content that not enough reflection and teacher expression time is built on. Teachers, just as students, are not empty receptacles that are to be filled with knowledge. A basic premise in this study, based on the principle of self-esteem, was for teacher trainers to avoid being transmitters of knowledge and instead strive to become mediators of thinking, to show respect for each teacher's contributions.

d. **Cooperative Learning for building communities of teachers.** The picture of teacher development that emerged from this study is in accordance with research on student's active learning. That is, students learn more effectively through participation in meaningful joint activities in which their performance is assisted by a more capable peer (Vygotsky, 1978; Tharp and Galimore, 1989; Duran, 1990).

It is also natural for adults to learn together and expand what Vygotsky called their zone of proximal development. As teachers worked together in cooperative teams, they developed a quicker understanding and transfer of the content. More important, they developed an ecology conducive to continued personal and professional growth. Bilingual mentor teachers reported that they had learned as much as the beginning teachers. Both partners were experts at something. Mentor teachers had the seasoned experiences of years of teaching and problem solving. Bilingual beginning teachers had current knowledge of new teaching strategies and approaches. Each one took turns being "more capable peer." This assisted performance built self-respect and respect for other colleagues.

Conclusion

The grant paper concentrated on identifying content for preparing teachers for LEP students. Chamot (1991); Garcia (1988). Calderon (1981; 1982, 1984, 1986, 1988, 1991) and others have identified similar content. However, we can see from the body of research and on-going recent projects, that the issues of teacher training need to be explored at a more macro level -- organizational structures at universities and schools - and at a micro level -- processes for preparing teachers to master the content necessary to effectively instruct LEP students.

The essentially social nature of teaching and learning needs to be emphasized in teacher preparation courses and staff development sessions. By participating in such interactions, sometimes as an equal member and sometimes as a coach, teachers can study, reflect, analyze, model, practice, critique, and explain how to engage in teaching in ways appropriate to LEP student learning.

Figure 1

SUMMARY OF OBSERVATIONS: STAGES OF IMPLEMENTATION OF COOPERATIVE LEARNING

STAGE 1	STAGE 2	STAGE 3
• Interest	• Loses Interest	• High Interest
• Traditional Learning Groups but no CL strategies	• Does some CL mostly The same 2 OR 3 strategies	• Uses 4-5 Cooperative Learning strategies
• Tradit. discipline, problems with noise or no interaction	• Tradit. discipline problems with noise and discipline	• Good discipline, noise level, movement
• Homogeneous groups only	• Experimenting with grouping & problems	• Heterogeneous grouping
• No group goal, no individ. accountability only task emphasized	• A group goal, task, Zero Noise Signal but no individual accountability	• Structures group goal, rules, roles, Zero Noise Signal, tasks
• Problems with grading	• Problems with grading	• System scoring/grading
• Low-level content Traditional Learning Groups activities, "rote" learning	• Low-level content for Trad. Learning Groups & Coop. L.Grps. "rote" learning	• Low level content for CLG activities "rote" learning
• Students work on own, or copy other students' work, one appointed leader	• Some students work on own or copy other students' work, one leader evident	• One or two students are not cooperating or are copying others
• Students work silently, they don't want to share	• Students work with little interaction	• Students discuss but approach most tasks individually
• Teacher talk predominates, little time on group work	• Teacher's directions are too long or too short	• Teacher's directions and time for group work are appropriate
• Little or no teacher monitoring of groups	• Monitors on-task and discipline problems	• Monitors on-task, noise, discipline, & clarifies
• Debriefing/processing does not occur	• Debriefing/processing does not occur	• Simple debriefing occurs frequently
• Concerned with own teaching, appraisal outcomes, reluctant to try something new	• Concerned with own teaching, of losing student control, of wasting time with CLG and TLG.	• Concerned with student learning, classroom management

© Margarita Calderón, Ph. D. UCSB/YSLETA I.S.D. 3001 Cabot, El Paso, Tx. 79935 Ph.(915)595-5971

Figure 1 (Continued)

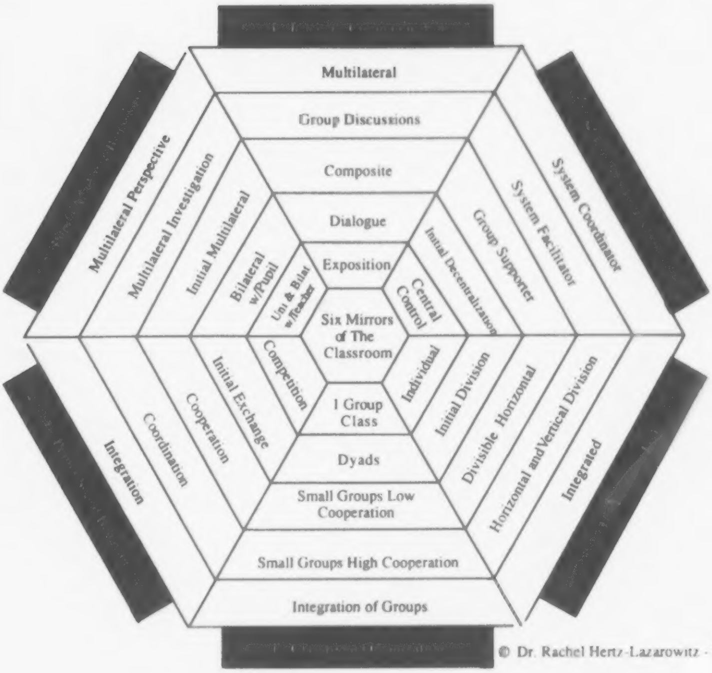
SUMMARY OF OBSERVATIONS: STAGES OF IMPLEMENTATION OF COOPERATIVE LEARNING

STAGE 4	STAGE 5
• High interest	• High interest
• Uses many CLG simple techniques and strategies	• Uses many CLG simple and complex techniques & strategies
• Can improvise a CLG lesson on the spur of the moment	• Can improvise any CLG lesson on the spur of the moment
• Uses a variety of simple and complex techniques & strategies	• Uses a variety of simple and complex techniques & strategies
• Carefully structures group goal, rules, roles, task, time, materials emphasizes indiv. accountability & responsibility for each other	• Carefully structures group goal, rules, roles, tasks, time, materials emphasizes indiv. accountability & responsibility for each other
• Great handle on discipline, reward structures	• Great handle on discipline, reward structures
• System for scoring/grading	• Integrated scoring/grading
• Emphasis on social skill development	• Emphasis on social skill, leadership skills, and creativity
• Low-level content for CLG activities	• High-level content for CLG activities
• Students help each other, reach consensus, shared leadership	• Students help each other, have negotiation process, make joint decisions on everything
• There is ample student interaction	• Ample multiple types of student interaction: communication, reasoning and scaffolding
• Teacher directions are abbreviated since students know the teaching models	• Teacher becomes facilitator of student organized learning, encourages self-reliance, choices
• Monitors, clarifies, takes notes for feedback	• Monitors, clarifies, provokes higher order thinking, facilitates
• Debriefing of content and process occurs systematically after each lesson	• Debriefing for higher order thinking occur systematically for content and process, for longer time
• Exhibits fidelity to the model, good pacing, control and smooth transitions	• Exhibits executive control of CLG, but there is flexibility and own adaptations work very well
• Concerned with adaptation to student needs & curriculum and smooth transitions	• Concerned with student outcomes, curriculum adaptation, training of other teachers

© Margarita Calderón, Ph. D. UCSB/YSLETA I.S.D. 3001 Cabot,
El Paso, Tx. 79935 Ph.(915)595-5971

BEST COPY AVAILABLE

Figure 2



© Dr. Rachel Hertz-Lazarowitz - '88-'92

947

References

Ada, A.F. (1968) Creative education for bilingual teachers. Harvard Educational Review, 56, (4), 386-394.

Atwell, R.H. (January 27, 1988). Ford and Carter to head panel on minorities, Interview with Robert Rothman for Education Week 8, (8)

Calderon, M., Hertz-Lazarowitz, R. & Tinajero, J. (in press) Bilingual students in CIRC: Moving from traditional and cooperative learning. In Williamson, J. (Ed.) Cooperative Learning: International Perspectives. London, England: Fulton Publishing Co.

Calderon, M. (in press) Mentoring and Coaching Minority Teachers. In De Villar & Cummins (Eds.) Successful cultural diversity: Classroom practices for the 21st century. New York: SUNY.

Calderon, M. (1991). Training strategies for mentors of minority teachers. In Texas Education Agency (Eds.) Handbook for Mentoring. Austin, TX: Texas Education Agency.

Calderon, M., Hertz-Lazarowitz, R. & Tinajero, J. (1991 October). Adapting CIRC to multiethnic and bilingual classrooms. Cooperative Learning Magazine. Santa Cruz, CA: The Magazine for Cooperative Learning. 12, (1).

Calderon, M., Hertz-Lazarowitz, R. & Tinajero, J. (in press) Bilingual students in CIRC: Moving from traditional to cooperative learning. In Williamson, J., (Ed.) Cooperative Learning: International Perspectives. London, England: Fulton Publishing Co.

Calderon, M., Tinajero, J. & Hertz-Lazarowitz (1991 winter). Adapting CIRC to meet the needs of bilingual students. Journal of Educational Issues of Linguistic Minority Students.

Calderon, M. & Blau, S. (1991). Teaching writing in a multicultural classroom. In S. Blau (Ed.) The writers craft: From idea to expression, Teachers' Edition. Evanston: McDougal, Little Publishers.

Calderon, M. (1990-91 Fall-Winter) Cooperative learning builds communities of teachers. In Journal of Teacher Education and Practice. 6, (2), 75-79.

Calderon, M., Tin ajero, J. & Hertz-Lazarowitz, R. (1990) Promoting student interaction for biliteracy development: Bilingual Cooperative Integrated Reading and Composition. In Research for the 1990s. McClaren-Lighty, M. (ed.) Santa Cruz, CA: University of California, Santa Cruz.

Calderon, M. (1990) Cooperative Learning for Limited English Proficient Students. Baltimore, MD: Johns Hopkins University.

Calderon, M. (1990) Cooperative Learning for Limited English Proficient Students. Austin, TX: Texas Education Agency, Distinguished Scholar Series Publication.

Calderon, M. (1990) Coaching and Collegial Collaboration. El Paso, TX: Ysleta Independent School District.

Calderon, M. (1989). Cooperative Learning for Limited English Proficient Students. IDRA Newsletter. San Antonio, TX: Intercultural Development Research Association.

Calderon, M. (1990). Cooperative Learning for Limited English Proficient Students. Austin, TX: Texas Education Agency, Distinguished Scholar Series Publication.

Calderon, M. (1989). Cooperative Learning for Limited English Proficient Students. IDRA Newsletter. San Antonio, TX: Intercultural Development Research Association, XVI, (9), 1-7.

Calderon, M. (1989). Sheltered instruction, Honolulu, HI: Hawaii State Department of Education.

Calderon, M. (1989). Applying research on effective bilingual instruction in a multi-district inservice teacher training program. National Association for Bilingual Education Journal, 12, (2), 133-152.

Calderon, M. (1988). Teaching Strategies Sections. In J. Tinajero et.al. "Stepping stones" MacMillan Transitional Reading Program, New York, NY.

Calderon, M. (1986). A trainer of trainers model: Focus on acquisition of literacy by Limited English Proficient students. Report #GXX8410030, Washington, DC: Secretary of Education.

Calderon, M. & Speigel-Coleman, S. (1985) Effective instruction for language minority students--from theory to practice. In Teacher Education Journal, 2, (3).

Calderon, M. (1984) Application of Innovation Configurations to a Trainer of Trainers Program, Bilingual Education Selected Papers Series, Los Angeles, CA: Evaluation Dissemination and Assessment Center.

Calderon, M. (1984) Training bilingual trainers: An ethnographic study of coaching and its impact on the transfer of training. doctoral dissertation, in Dissertation Abstracts and Claremont Graduate School/San Diego State University, 1984.

Calderon, M. (1984). Second language acquisition and organization change through an effective training model, Bilingual Education Selected Paper Series, Los Angeles, CA: Evaluation, Dissemination and Assessment Center.

Calderon, M. & Cummins. (1982). Communicative competence in bilingual education: Theory and research. Dallas, TX: National Center for the Development of Bilingual Curriculum.

Calderon, M., Di Pietro, R. & Rubio, O. (1981). Methods and techniques for communicative competence. Dallas, TX: National Center for the Development of Bilingual Curriculum.

Calderon, M., Cummings, J. & Pusey, M. (1981). Assessing communicative competence. Dallas, TX: National Center for the Development of Bilingual Curriculum.

Calderon, M. & Belker, J. (1981). Management system for the bilingual education teacher training series. Dallas, TX: National Center for the Development of Bilingual Curriculum.

Clark, E. and Milk, R. (1983). Training bilingual teachers: A look at the Title VII graduate in the field. National Association for Bilingual Educators Journal, 8, (1), 42-53.

Cummins, J. (1989). Empowering minority students. California: California Association for Bilingual Education.

Hertz-Lazarowitz, R. (in press). Understanding students' interactive behavior: Looking at six mirrors of the classroom. In R. Hertz-Lazarowitz & Miller, N. (Eds.), Interaction in cooperative groups: The anatomy of group learning. New York: Cambridge press.

Houston, W.R. & Calderon, M. (1990-91). Preparing an ethnically diverse teaching force. In Cuellar & Huling-Austin (Eds.). Achieving an Ethnically-Diverse Teaching Force. Austin, Tx: Texas Education Agency.

Houston, W.R. & Calderon, M. (1990-91 Fall/Winter). Preparing an ethnically diverse teaching force. In Journal of Teacher Education and Practice. 6, (2), 43-48.

Mata, S et al, (1987). Improvement of pre-service teacher education through clinical supervision of student teachers: Supervision and support through the eyes of student teachers and first year teachers. San Francisco, CA: Far West Laboratory.

Mata, S. (in press). Options for increasing the pool of bilingual teachers. In Options for meeting the bilingual teacher demand. Sacramento: California State Department of Education.

Sharan, S., Kusseu, P., Hertz-Lazarowitz, R. Bejarano, Y., Raviv, S. & Sharan, Y. (1984). Cooperative Learning in the classroom: Research in desegregated schools. Hillsdale, NJ: Erlbaum.

Slavin, R.E. (1983). Cooperative learning. New York: Longman.

Slavin, R.E. (1986). Using student team learning. Baltimore, MD: Center for Research on Elementary and Middle School, Johns Hopkins University.

Slavin, R.E. (1987 October). Cooperative Learning and the Cooperative School. Educational Leadership Journal. 45, (2), 11-23.

Tharp, R.G. and R. Gallimore (1989). Rousing Minds to Life: Teaching, Learning, and Schooling in Social Context. Cambridge: Cambridge University Press.

Tinajero, J. & Calderon, M. (1988) Language approach plus. Journal of Educational Issues of Language Minority Students, 1 , (1), 1988.

Vygotsky, L.S. (1978). Mind in Society. Cambridge, MA: Harvard University Press.

Response to Carl Grant's Presentation

Li-Rong Lilly Cheng
San Diego State University

It gives me great pleasure to attend this Second National Symposium on Limited English Proficient Students Issues sponsored by the Office of Bilingual Education and Minority Languages Affairs. The research and theoretical speculations presented, particularly the paper by Dr. Carl Grant entitled Successful Innovations in Teacher Education Programs gave me occasion to reexamine some fundamental concerns regarding responsibility, dominant paradigms and potential solutions to the LEP situation. My conclusions intimate a need for nothing less than a radical shift in focus and essence in all academic fields: research, training, curricula and, at heart, the very definitions of knowledge and culture as transmitted to our students. My perspective as both an outsider looking in (an LEP first generation immigrant), and an insider looking out (a scholar of the immigrant/LEP phenomenon) leads me to believe that our current methods and terms for dealing with the LEP student are in need of scrutinization and restructuring.

In examining the term "limited English proficiency," one might detect a type of cultural bias; the very word 'limited' connotes a type of reproach or judgment, that a student is not "proficient" in English is labeled negatively. Because the student does not speak the language of the teacher; the communication breakdown, the mixed signals, and the gradual marginalization process may result in irreparable educational and psychological/emotional damage. If a child arrives from Laos who is Hmong, neither he nor his parents speak English, his teacher cannot speak Hmong....the typical LEP problem is presented. How can we empower the student instead of making him feel limited? How can the teacher react besides growing frustrated and ignoring the problem?

The instructor might undertake some responsibility by altering the "school discourse" to encourage bilingual development, intensive training in English, examining interactions, attempting to relate to and identify with the student, creating a compassionate and positive learning environment. Current instructional patterns may not allow room for learning about, understanding, and respecting the cultures of students from diverse background. The following model (Cheng, 1990) suggests ways in which traditional paradigms might shift in emphasis to better facilitate a culturally diverse classroom. The Existing Model column on the left lists regular methods of coping with LEP, the corresponding column on the right provides an alternative pedagogical/psychological technique. For example, when our Hmong student arrives on his first day in the American classroom, instead of viewing his difference as a deficit, his teacher could encourage it as an asset, his shaky English might be "enhanced" and improved on, instead of compensated for. In short, LEP students will feel less "limited" if their difference is labeled positive rather than negative.

Existing Model	Paradigm Shift
Compensatory	Enhancement
Reduction	Addition
Standard	Diverse
Assimilation	Multiculturalism
Deficit	Asset
Tolerance	Acceptance
Disenfranchise	Empower

In the body of research on LEP students, not much attention is paid to school discourse, namely; the interactions between teachers and students, students and students and schools and family/community. Dr. Carl Grant, in his paper, mentions the scarcity of research in this important area. Again, the responsibility falls to the teachers. Since they are the individuals who are best qualified to gather information and make assessments, since they work most intimately with the students, they should likewise provide the majority of related research data. The ongoing dialogue between teacher and student is our richest, most accessible, and, in my opinion, most important source of information for examining the matrix in which LEP problems originate. With a shift in emphasis to the nature of the student's experience, their relationship to dual/multiple cultures and their linguistic and sociological anxieties, we can develop a new research that provides the most revealing insights and, consequently, the potential for the most viable solutions.

The insights teachers could generate might illumine another unexamined problem of current LEP research -- that of the "hidden interaction." Most researchers base their studies on what can be deemed "explicit" indicators: grades, test scores, language proficiency, teacher evaluation, etcetera, seldom taking into account the "implicit" or hidden indicators of cultural/historical background, ethics, social codes, body language and so on, as they operate in the world of both the student and the teacher. Educators often assume that the beliefs and values, the rules and norms that traditionally dictate the decorum of an American classroom, are part and parcel of every student's experience. Certainly, this is not the case and these ill-founded assumptions are the cause of many misinterpretations. For example: a teacher comes into the classroom and says "Good morning," expecting the children to reply in turn. Because one student's culture has taught him to defer to elders, because he considers it disrespectful to answer one's teacher informally, he does not reply. His behavior has violated the teacher's "hidden agenda" and will most likely be construed as rude or labeled out of keeping with the "proper" social codes.

The implicit agenda also applies to assimilation and mainstreaming. Students who are misinterpreted and disenfranchised in a gradual process of conflicting messages, such as the one shown in the example above, are never given the passport to enter the mainstream culture. While explicitly we state that proficiency in English is enough to admit any individual, the rules by which American culture, and, in particular, American education, accept people are tacit and unspoken. These rules are never taught but are nevertheless in constant and rigorous effect. While a mastery of the complex systems of codes and attitudes that implicitly guide our interactions will never be attainable for the LEP student without "explicit" assistance and understanding, it still unfortunately remains the true passport to assimilation.

There was much talk at the symposia about sense-making -- how we make sense, the politics of communication and how we negotiate meaning. The challenge then is to shift our definitions of "meaning," to relocate our research in terms of implicit phenomena as opposed to explicit, to bring to light our cultural/behavioral agenda as it relates to the LEP problem. Teachers and students must negotiate between each other, themselves and their backgrounds/cultures in an attempt to achieve a new "sense" of their situation. I advocate the notion of "teacher-researcher" and would like to urge that data be collected by videotaping interactions, such a research method could prove a very powerful tool from which we can glean at least some of the hidden agenda: how students are interpreted, how information is exchanged, physical signals, dialogue analysis, etcetera. From a more extensive and more enlightened research, we can begin to develop pedagogical policies, much like the paradigm shifts mentioned earlier, that will lead us into more discussion and intervention tactics on how to empower LEP students.

Another barrier, also mentioned in Dr. Carl Grant's paper, is the issue of teacher education and its underlying philosophical and ethical tenets. In a large survey that Dr. Carl Grant quoted, teachers in training found many of their instructors, who advocated an active multicultural and multilingual classroom, seldom practiced what they preached. Whether students are in preservice or in-service training, whether they are in actual education fields or in fields such as history, communicative disorders, linguistics or anthropology, we have to look at the teachers, professors, and faculty who are passing on the "hidden agenda" to the next generation of teachers. If we are to train teachers who will make a difference, we need to examine our existing faculty, seeking not only those who endorse the notion of diversity but also those who can translate it into a practical and applicable method their students can understand and use.

In conclusion I would like to reiterate that the philosophical basis of our classroom interactions, the implicit messages and assumptions that escape our research and the planning of our curricula, the passport to assimilation whose requirements go far beyond mere language proficiency can and must be changed, shifted and expanded. Paradigms of culture and language are not objective or fixed, they are created and applied according to the ideals of human beings, they are malleable, open to revision and restructuring. Knowledge is so much more than the transmission of information, it is the learned code of life-skills such as critical thinking and problem solving. These abstractions are no longer terms of lofty nobility, they are the seeds of practical necessity, seeds we must plant collaboratively with our children. If we let our children lead us, if we allow them to lead us, we may find new solutions to the problems of "implicit interaction." If we open up a space in the classroom for them, if we educate our teachers to encourage the strengths and participation of LEP students and lastly, if we enlarge our conceptions of each other, increasing our sensitivities to include the experience of others -- more than half the battle has already been won.

INDEX

Volume II

RESEARCH PAPERS

Baker, Eva L. 1
Canales, JoAnn 89
Geisinger, Kurt F. 33
Grant, Carl A. 431
Haladyna, Thomas 135
Hamp-Lyons, Liz 317
Negroni, Peter J. 371
Rosebury, Ann S. 273
Secada, Walter 209
Steffens, John 393
Warren, Beth 273

DISCUSSANTS

Byron, Peter M. 27
Calderon, Margarita 457
Cheng, Li-Rong Lilly 473
Collier, Virginia 417
Galang, Rosita 423
Hargett, Gary 165
Hewlett-Gomez, Michele R. 69
Lara, Julia 123
Lindquist, Mary 267
Martinez, Robert 199
McKeon, Denise 359
Pennock-Román, Maria 173
Peterson, Penelope L. 257
Peyton, Joy Kreeft 365
Pierce, Lorraine Valdez 19
Rohac, Ron 305
Rudner, Lawrence M. 81
Rueda, Robert 127
Tsang, Sau-Lim 311

PANELISTS

Balu, Raj 183
Berney, Tomi Deutshc 193
Salazar, Jesus 187